IRIS AperTO

UNIVERSITÀ DEGLI STUDI DI TORINO

AperTO - Archivio Istituzionale Open Access dell'Università di Torino

An analytical study of participating policies with minimum rate guarantee and surrender option

This is the author's manuscript

Original Citation:

Availability:

This version is available http://hdl.handle.net/2318/1879522 since 2022-11-14T22:53:00Z

Published version:

DOI:10.1007/s00780-022-00471-0

Terms of use:

Open Access

Anyone can freely access the full text of works made available as "Open Access". Works made available under a Creative Commons license can be used according to the terms and conditions of said license. Use of all other works requires consent of the right holder (author or publisher) if not exempted from copyright protection by the applicable law.

(Article begins on next page)

12 October 2024

AN ANALYTICAL STUDY OF PARTICIPATING POLICIES WITH MINIMUM RATE GUARANTEE AND SURRENDER OPTION

MARIA B. CHIAROLLA, TIZIANO DE ANGELIS, AND GABRIELE STABILE

Abstract. We perform a detailed theoretical study of the value of a class of participating policies with four key features: (*i*) the policyholder is guaranteed a minimum interest rate on the policy reserve; (*ii*) the contract can be terminated by the holder at any time until maturity (surrender option); (*iii*) at the maturity (or upon surrender) a bonus can be credited to the holder if the portfolio backing the policy outperforms the current policy reserve; (*iv*) due to solvency requirements the contract ends if the value of the underlying portfolio of assets falls below the policy reserve.

Our analysis is probabilistic and it relies on optimal stopping and free boundary theory. We find a structure of the optimal surrender strategy which was undetected by previous (mostly numerical) studies on the same topic. Optimal surrender of the contract is triggered by two 'stop-loss' boundaries and by a 'too-good-to-persist' boundary (in the language of [17]). Financial implications of this strategy are discussed in detail and supported by extensive numerical experiments.

1. Introduction

Participating Policies with Minimum Rate guarantee are insurance contracts, appealing predominantly to individuals during their working lives, as a form of low-risk financial investment. The subscriber of a participating policy (policyholder) pays a premium (either single or periodic) which is used by the insurance company to set up a so-called policy reserve for the policyholder. The policy reserve is linked to a portfolio of assets held by the insurance company and it accrues interest tracking the performance of such portfolio (the details of the contract are illustrated in Section 2). The minimum rate guarantee is in the form of a minimum interest rate paid by the insurance company towards the policy reserve irrespective of the performance of the portfolio backing the policy (this rate is usually lower than the risk-free rate). In the absence of any further contract specifications, the policy terminates at a given maturity, at which the policyholder receives an amount equal to the value of the reserve, plus a *bonus*, if the current value of the portfolio is sufficiently high relative to the reserve.

The contract may incur early termination. That happens if the value of the portfolio backing the policy is not sufficient to cover the policy reserve. In that case we say that the insurance company fails to meet the solvency requirements on the participating policy and the policyholder receives the value of the policy reserve at that time. More interestingly, early termination of the contract may be an *embedded option* in the policy specification. Indeed, along with the standard contract, the policyholder can buy the right to an early cancellation of the policy, the so-called *surrender option* (SO). If the

Date: November 11, 2021.

2010 *Mathematics Subject Classification.* 91G80, 62P05, 60G40, 35R35; *JEL Classification.* G22.

Key words and phrases. Participating policies, minimum rate guarantee, surrender option, solvency requirement, optimal stopping, free-boundary problems, stop-loss boundary, too-good-to-persist boundary.

policyholder exercises the surrender option prior to the maturity of the contract, at that time she receives the value of the policy reserve, plus the above mentioned *bonus*.

While surrender options share similarities with financial options of American type, due to their early exercise feature, they are actually rather different in nature. Indeed, the presence of a surrender option, embedded in a participating policy, changes the structure of the whole contract. As a consequence, the price of embedded options is normally defined as the difference between the value of a policy which includes the option and the value of a policy which does not include the option (see (2.9) for a mathematical expression).

Participating Policies with Surrender Option (PPSO) have been studied extensively in the academic literature. This paper contributes to that strand of the literature which assumes that the policyholder is fully rational and the surrender option is exercised optimally from a financial perspective. Other papers analyse PPSO in which surrender occurs as a randomised event (see Cheng and Li [9]) or without assuming rational behaviour of the policyholder (see Nolte and Schneider [29]). Several papers adopt a numerical approach to analyse PPSO without solvency requirements for the insurance company (see Andreatta and Corradin [1], Bacinello [2], Bacinello, Biffis and Millossovich [3], Grosen and Jørgensen [22] among others). Chu and Kwok [10] provide an analytical approximation for the price of a participating policy without taking into account the surrender option. Finally, Siu [34] considers the fair valuation of a PPSO when the market value of the portfolio backing the policy is modelled by a Markov-modulated Geometric Brownian Motion. In [34] the author approximates the solution of a free boundary problem by a system of second-order piecewise linear ordinary differential equations.

In this paper we develop a fully theoretical analysis of participating policies with minimum rate guarantee, embedded surrender option and early termination due to solvency requirements. Following the example of other papers on this topic (see, e.g., Chu and Kwok [10], Fard and Siu [18], Grosen and Jørgensen [22], Siu [34]), we focus purely on the financial aspects of the policy and ignore the demographic risk, in the sense that the contract does not account for a possible demise of the policyholder. From the point of view of applications we may imagine that there are multiple beneficiaries of the policy, so that the demographic risk is negligible. Moreover, it is well-known (see, e.g., Cheng and Li [9], Stabile [35]) that the assumption of a constant force of mortality, independent of the financial market, results in a shift in the discount rate adopted for pricing; this does not affect the methods we employ and the qualitative outcomes of our work. The study becomes substantially more involved if one considers a time-dependent mortality force (as, e.g., in De Angelis and Stabile [15]) or worse a stochastic mortality. We leave these extensions for future work.

Our main contributions are: (i) the analytical study of the pricing formula for the PPSO, (ii) a characterisation of the optimal exercise strategy for the surrender option, in terms of an optimal exercise boundary, and (iii) an extensive numerical analysis of the option value and the surrender boundary. The arbitrage-free price of the policy is obtained as the value function of a suitable finite-time horizon optimal stopping problem, on a two-dimensional degenerate diffusion which lives in an orthant of the plane and is absorbed upon leaving the orthant. The diffusive coordinate of such process models the dynamics of the portfolio backing the policy, while the other coordinate represents the dynamics of the policy reserve. After a suitable transformation, the dynamics is reduced to a one dimensional diffusion in the form of a stochastic differential equation (SDE) with absorption upon hitting zero. This process corresponds to the so-called bonus distribution rate of the contract, which will be introduced in more detail in the

next section. We are then led to consider a state process $(t, X_t) \in [0, T] \times \mathbb{R}_+$ which is absorbed if X_t reaches zero, so that our state space is the (t, x)-strip with $t \in [0, T]$ and $x \geq 0$.

The optimal stopping problem poses a number of challenges: it is set on a finite-time horizon, hence it is not amenable to explicit solutions using the associated free boundary problem (which is indeed parabolic); the SDE that describes the stochastic process does not admit an explicit solution, so that numerous tricks often used in optimal stopping problems, and relying on an explicit dependence of the process on its initial value, are not applicable (see, e.g., the American put problem in Peskir and Shiryaev [30]); the stopping payoff is independent of time but, as a function of x, it is convex with discontinuous first derivative.

The combination of the above ingredients produces a very peculiar shape of the optimal stopping region, which we derive from a detailed analysis of the value function. We observe that the stopping region $\mathcal{S}$, i.e., the points (t, x) at which the policyholder should instantly surrender the contract, is not connected in the x-variable, for each value of t given and fixed. Instead, $\mathcal{S}$ may have two connected components for each $t \in [0, T]$ (see Figure 1), corresponding to two distinct 'stop-loss' boundaries and a 'too-good-to-persist' boundary (following [17], here we say that a boundary is 'stop-loss' or 'too-good-to-persist' if $\mathcal{S}$ can be locally represented as a set $\{(t, x) : x \leq b(t)\}$ or $\{(t, x) : x \geq b(t)\}$, respectively, for some b). This result was not observed in prior work on the same model, where a numerical approach to the problem could not detect this unusual feature (see, e.g., Siu [34]). As it turns out, the shape of the stopping set is closely related to the bonus mechanism included in the PPSO and it has fine implications on the optimal exercise of the surrender option. We will elaborate further on this point in Section 3.3, once the mathematical details have been laid out more clearly.

The stopping set is connected in the t-variable, for each given $x \geq 0$. This leads naturally to consider an optimal stopping boundary as a function of x (rather than as a function of t, as in the vast majority of papers in the area). We obtain a wealth of fine properties of the map $x \mapsto c(x)$ on $[0, \infty)$, which are of independent mathematical interest for the probabilistic theory of free boundary problems. Indeed we show that $c(\,\cdot\,)$ is continuous on $(0, \infty)$ and piecewise monotonic, with two *strictly increasing* portions and a *strictly decreasing* one. It is important to remark that questions of continuity of the optimal boundary $x \mapsto c(x)$ are much harder to address than in the more canonical setting of time dependent boundaries $t \mapsto b(t)$. Here we resolve the issue in Theorem 5.11, by providing a probabilistic proof which is new in the literature and makes use of suitably constructed reflecting diffusions. Our proof provides a conceptually simple way to show (in more general examples) that time dependent optimal boundaries $t \mapsto b(t)$ cannot exhibit flat stretches, unless the smooth-fit property fails.

The rest of the paper is organised as follows. In Section 2 we set up the model in a rigorous mathematical framework and then we state our main results in Section 3 (Theorems 3.1 and 3.2). In particular, in Section 3.3 we obtain numerical illustrations of the value function, the stopping set and the related sensitivity analysis, accompanied by a financial interpretation. Section 4 contains preliminary technical results on the continuity and monotonicity of the value function. In Section 5 we analyse in detail the free boundary problem associated with the PPSO and we prove Theorems 3.1 and 3.2 stated in Section 3. Section 6 extends our framework to include management fees in the valuation of the PPSO. The paper is completed by a short technical appendix.

2. Actuarial model and problem formulation

In this section we provide a mathematical description of the price of a PPSO in a complete market, under a risk-neutral probability measure. We align our setup and part of our notations to those already used in other papers on this topic as, e.g., Chu and Kwok [10], Grosen and Jørgensen [22] and Siu [34].

Given $T > 0$, we consider a market with finite time horizon $[0, T]$ on a complete probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,T]}, \mathsf{Q})$ that carries a one-dimensional Brownian motion $\widetilde{W} := (\widetilde{W}_t)_{t\in[0,T]}$. With no loss of generality we assume that the filtration $(\mathcal{F}_t)_{t\in[0,T]}$ is generated by the Brownian motion and it is completed with Q-null sets. Our market is complete and Q is the risk-neutral probability measure.

An investor can purchase a PPSO at time zero by making a lump payment V_0 to an insurance company. In return, the insurer invests an amount R_0 into a financial portfolio and commits the company to credit interests to the policyholder's *policy reserve* according to a mechanism that will be described below. Thanks to the *surrender option* embedded in the contract, the policyholder has the right to withdraw her investment at any time prior to the policy's maturity T. In this case, she receives the so-called *intrinsic value* of the policy.

2.1. The policy reserve.

First we describe the rate at which the amount R_0 invested by the insurer accrues interest, based on the performance of the portfolio backing the policy (the *reference portfolio*). We let $A := (A_t)_{t\in[0,T]}$ be the process denoting the value of such portfolio and assume that A evolves as a geometric Brownian motion under Q; that is,

$$\begin{cases} \mathrm{d}A_t = A_t\big(r\mathrm{d}t + \sigma \mathrm{d}\widetilde{W}_t\big), \\ A_0 = a_0, \end{cases} \tag{2.1}$$

where r, σ and a_0 are positive constants and r is the risk free rate.

During the lifetime of the policy, the *policy reserve* is denoted by $R := (R_t)_{t\in[0,T]}$ and accrues interest based on a two-layer mechanism. First, the insurance company guarantees a minimum fixed interest rate, which we denote by r^G and, in line with financial practice, we assume

$$r^G \in (0, r). \tag{2.2}$$

Second, at times when the portfolio performs particularly well, the policyholder participates in the returns. In particular, we define the so-called *bonus reserve* $B_t := A_t - R_t$ and, as in [10], [34], [33] and [18], we consider a *bonus distribution rate* (BDR) of the form

$$\ln\big(1 + B_t/R_t\big) = \ln\big(A_t/R_t\big). \tag{2.3}$$

The BDR measures the performance of the portfolio against the performance of the policy reserve. The insurance company compares the BDR to a constant, long-term target $\beta > 0$, known as *target buffer ratio*. If the BDR exceeds the target buffer ratio, a proportion $\delta > 0$ of the excess is shared with the policyholder.

Combining the minimum guaranteed interest rate with the bonus rate gives the instantaneous rate of interest on the policy reserve, that is

$$c(A_t, R_t) = \delta\bigg(\ln\Big(\frac{A_t}{R_t}\Big) - \beta\bigg) \vee r^G. \tag{2.4}$$

It follows that the policy reserve R evolves under Q according to the dynamics

$$\begin{cases} \mathrm{d}R_t = c(A_t, R_t)R_t\mathrm{d}t, \\ R_0 = \alpha\, a_0, \end{cases} \tag{2.5}$$

where $\alpha \in (0,1)$ is fixed by the insurer. Hence, the initial reserve R_0 covers α shares of the reference portfolio.

Remark 2.1. *Notice that in the specification of the bonus mechanism in (2.4) we may equivalently consider* $\ln(\alpha A_t/R_t)$ *instead of* $\ln(A_t/R_t)$*. This would emphasise that the policyholder only receives a bonus proportional to her share of the portfolio backing the policy. From the mathematical point of view, of course there is no difference since* $\ln(\alpha A_t/R_t) = \ln\alpha + \ln(A_t/R_t)$ *and the additional term,* $\ln\alpha$*, is absorbed in the specification of the target buffer ratio* $\beta > 0$*.*

2.2. **Intrinsic value of the policy and arbitrage-free price.** Next we describe the so-called intrinsic value of the policy, which is the value that the policyholder receives either at the maturity of the policy or at an earlier time, should she decide to exercise the surrender option.

The intrinsic value is equal to the policy reserve plus a bonus component. The latter, is activated when the value of the policyholder's shares in the portfolio A exceeds the current value of the policy reserve; that is, when $\alpha A_t > R_t$. In this case the policyholder receives a bonus fraction γ of the surplus of her α-share.

From the mathematical point of view, the intrinsic value of the policy may be written as

$$g(A_t, R_t) := R_t + \gamma\,[\alpha A_t - R_t]^+\,, \tag{2.6}$$

where $[x]^+ := \max\{x, 0\}$ and $\gamma \in (0,1)$ is the so-called *participation coefficient*.

The model also takes into account that the company may fail to meet the solvency requirement at any time before T. In fact, denoting by $\tau^\dagger$ the stopping time (*insolvency time*)

$$\tau^\dagger := \inf\{t \geq 0 \,:\, A_t \leq R_t\}, \tag{2.7}$$

the company's solvency requirement is satisfied for $t < \tau^\dagger$. In the event of $\tau^\dagger < T$ the policy is liquidated and the policyholder receives (cf. (2.6))

$$g(A_{\tau^\dagger}, R_{\tau^\dagger}) = R_{\tau^\dagger},$$

i.e., the policy reserve value.

Finally, we can define V_0, the arbitrage-free price of the PPSO at time zero. Notice that $V_0 = V_0(\alpha)$, in the sense that the contract is specified by indicating the portion α of the portfolio which backs the policy. Recalling (2.1), (2.5), (2.6) and (2.7), we have

$$V_0 = \sup_{0\leq\tau\leq T} \mathsf{E}^{\mathsf{Q}}\Big[e^{-r\,(\tau\wedge\tau^\dagger)} g(A_{\tau\wedge\tau^\dagger}, R_{\tau\wedge\tau^\dagger})\Big], \tag{2.8}$$

where E^{Q} is the expectation under the measure Q and the supremum is taken over all stopping times $\tau \in [0,T]$ with respect to $(\mathcal{F}_t)_{t\in[0,T]}$. In what follows we will refer to (2.8) as the *PPSO problem*.

The value of the surrender option embedded in the contract (usually referred to as Early Exercise Premium in the mathematical finance literature) can be obtained as

$$V^{\mathrm{opt}} := V_0 - V_0^E, \tag{2.9}$$

where V_0^E is the arbitrage-free price of the contract without the possibility of an early surrender, that is

$$V_0^E = \mathsf{E}^{\mathsf{Q}}\Big[e^{-r\,(T\wedge\tau^\dagger)} g(A_{T\wedge\tau^\dagger}, R_{T\wedge\tau^\dagger})\Big]. \tag{2.10}$$

It is worth noticing that, in practice, the use of surrender options may be disincentivised by insurance companies, who agree to pay out only a fraction of the policy reserve in case of early surrender. In our case that would correspond to take $\lambda g(A_\tau, R_\tau)$ in (2.8) on the event $\{\tau < \tau^\dagger \wedge T\}$, with $\lambda \in [0,1)$. Here we focus on the model set out in (2.8), i.e., $\lambda = 1$, which is consistent with the existing literature (see, e.g., [22], [34]) and provides an upper bound for the prices of contracts with $\lambda \in [0,1)$. As we will see below, this model is also the source of interesting mathematical findings from the point of view of optimal stopping theory.

2.3. **Dimension reduction and bonus distribution rate.** As noticed in [10] and [34], the PPSO problem can be made more tractable by considering the bonus distribution rate (2.3) (i.e., the logarithm of the ratio A/R) as the observable process in the optimal stopping formulation of (2.8). Indeed, set $X := (X_t)_{t\in[0,T]}$ with

$$X_t := \ln\left(\frac{A_t}{R_t}\right) \qquad \text{for } t \in [0,T],\ \mathsf{Q}\text{-a.s.} \tag{2.11}$$

Then, by (2.1) and (2.5), one gets

$$\mathrm{d}X_t = \Big(r - r^G - \tfrac{1}{2}\sigma^2 - \big[\delta(X_t - \beta) - r^G\big]^+\Big)\mathrm{d}t + \sigma \mathrm{d}\widetilde{W}_t, \tag{2.12}$$

with initial condition

$$X_0 = x_\alpha := \ln(1/\alpha) > 0. \tag{2.13}$$

In terms of X the insolvency time becomes

$$\tau^\dagger = \inf\{t \geq 0 \,:\, X_t \leq 0\}. \tag{2.14}$$

Next we write the intrinsic value of the policy $g(A_t, R_t)$ (see (2.5)) in terms of X_t. For $x \in \mathbb{R}_+ := [0,\infty)$ define the gain function

$$h(x) := e^{-x} + \gamma\big[\alpha - e^{-x}\big]^+, \tag{2.15}$$

and notice that

$$x \mapsto h(x) \text{ is convex, striclty decreasing and } |h(x) - h(y)| \leq |x - y|, \tag{2.16}$$

since $\gamma \in (0,1)$. Then

$$g(A_t, R_t) = A_t\Big(e^{-X_t} + \gamma\big[\alpha - e^{-X_t}\big]^+\Big) = A_t\, h(X_t), \qquad t \in [0,T],\ \mathsf{Q}\text{-a.s.} \tag{2.17}$$

From the expression above we notice that for $X_t > x_\alpha$ the participation bonus in the intrinsic value of the policy is strictly positive. So we can think of x_α as the activation threshold for the participation bonus.

Now the key to the dimension reduction is a change of measure. Define the martingale process $M := (M_t)_{t\in[0,T]}$ by

$$M_t := e^{\sigma\widetilde{W}_t - \frac{\sigma^2}{2}t} = e^{-rt} A_t/a_0, \tag{2.18}$$

and the probability measure P equivalent to Q on $\mathcal{F}_T$ given by $\mathrm{d}\mathsf{P} = M_T\,\mathrm{d}\mathsf{Q}$. By Girsanov theorem the process $W := (W_t)_{t\in[0,T]}$ with

$$W_t := \widetilde{W}_t - \sigma t, \tag{2.19}$$

is a P-Brownian motion. Then, under the new measure P, the dynamics of X reads

$$\begin{cases} \mathrm{d}X_t = \pi(X_t)\mathrm{d}t + \sigma \mathrm{d}W_t, \\ X_0 = x_\alpha > 0, \end{cases} \tag{2.20}$$

where

$$\pi(x) := r - r^G + \frac{1}{2}\sigma^2 - \big[\delta(x-\beta) - r^G\big]^+. \tag{2.21}$$

For future reference, it is worth defining

$$\bar{x}_0 := \beta + \frac{r}{\delta} \quad \text{and} \quad x_G := \beta + \frac{r^G}{\delta}. \tag{2.22}$$

Then $x_G < \bar{x}_0$ since $r^G < r$. When the BDR process exceeds x_G (i.e., $X_t > x_G$) the policyholder receives the bonus interest rate on the reserve, above the minimum rate guaranteed r^G (see (2.4)). Again by (2.4) the interest rate paid on the reserve is higher than the risk-free rate r when the BDR process exceeds $\bar{x}_0$ (i.e., $X_t > \bar{x}_0$).

Using (2.17), (2.18) and the optional sampling theorem, for any stopping time $\tau \in [0,T]$ one easily obtains

$$\mathsf{E}^{\mathsf{Q}}\Big[e^{-r\,(\tau\wedge\tau^\dagger)} g(A_{\tau\wedge\tau^\dagger}, R_{\tau\wedge\tau^\dagger})\Big] = a_0\, \mathsf{E}^{\mathsf{Q}}\Big[M_{\tau\wedge\tau^\dagger}\, h\big(X_{\tau\wedge\tau^\dagger}\big)\Big] = a_0\, \mathsf{E}\Big[h\big(X_{\tau\wedge\tau^\dagger}\big)\Big], \tag{2.23}$$

with $\mathsf{E}[\,\cdot\,]$ denoting the P-expectation. Hence, (2.8) may be rewritten as $V_0 = a_0 v_0$, where

$$v_0 := \sup_{0\le\tau\le T} \mathsf{E}\big[h\big(X_{\tau\wedge\tau^\dagger}\big)\big]. \tag{2.24}$$

Life-insurance contracts often charge management fees to the holder. Methods developed in this paper can be used to deal with fees that are charged at a rate proportional to the value of the portfolio A and to the policy reserve R. If we add proportional fees to our model, the arbitrage-free price of the PPSO changes from its expression in (2.8) to

$$V_0 = \sup_{0\le\tau\le T} \mathsf{E}^{\mathsf{Q}}\Big[e^{-r\,(\tau\wedge\tau^\dagger)} g(A_{\tau\wedge\tau^\dagger}, R_{\tau\wedge\tau^\dagger}) - \int_0^{\tau\wedge\tau^\dagger} e^{-rs}\Big(pA_s + qR_s\Big)\mathrm{d}s\Big], \tag{2.25}$$

for some $p, q \ge 0$ (notice in particular that $p = p(\alpha)$ should depend on the fraction α of the portfolio backing the policy purchased by the policyholder). Also in this case we can perform the change of measure displayed above and arrive at $V_0 = a_0 v_0$, where now

$$v_0 = \sup_{0\le\tau\le T} \mathsf{E}\Big[h\big(X_{\tau\wedge\tau^\dagger}\big) - \int_0^{\tau\wedge\tau^\dagger} \Big(p + qe^{-X_s}\Big)\mathrm{d}s\Big]. \tag{2.26}$$

It is important to emphasise that, given a participation level $\alpha > 0$, the value V_0 of the PPSO is specific to such value of α. Indeed both the initial value of the reserve R_0 and the participation bonus on the intrinsic value of the policy depend on α. So we should think of the PPSO's value as $V_0 = V_0(\alpha)$ (and equivalently $v_0 = v_0(\alpha)$). In order to solve the problem, i.e., determine V_0 and the optimal exercise time of the surrender option, in the next section we will embed our problem in a Markovian setting.

3. Summary of main results

Thanks to the Markovian nature of the process X the value v_0 only depends on the initial value of the process $X_0 = x_\alpha$ (see (2.13)) and on the maturity T of the contract. However, in order to be able to characterise v_0 and the associated optimal stopping rule, we must embed our problem into a larger state-space by considering all possible initial values of the time-space dynamics (t, X). For that we denote by X^x the process

X starting at time 0 from an arbitrary point $x \geq 0$ and evolving according to (2.20). Similarly, we denote by $X^{t,x}$ the process X starting at time t from $x \geq 0$ and evolving according to (2.20). For future reference, when notationally convenient we will use $\mathsf{E}_x[\,\cdot\,] := \mathsf{E}[\,\cdot\,|X_0 = x]$. Since X is time-homogeneous, it holds

$$\mathsf{Law}\big((s, X^{t,x}_s)_{s\geq t}\big) = \mathsf{Law}\big((t+s, X^x_s)_{s\geq 0}\big).$$

Then we can identify the dynamics $(s, X^{t,x}_s)_{s\in[t,T]}$ and $(t+s, X^x_s)_{s\in[0,T-t]}$, and use the latter in the problem formulation below. Thanks to time-homogeneity, we also have that $\tau^\dagger$ defined in (2.14) is independent of time. Sometimes we use $\tau^\dagger(x)$ to emphasise that $\tau^\dagger$ depends on $X_0 = x$.

From now on we will study the finite-time horizon optimal stopping problem given by

$$v(t,x) := \sup_{0\leq\tau\leq T-t} \mathsf{E}\Big[h(X^x_{\tau\wedge\tau^\dagger})\Big], \qquad (t,x) \in [0,T]\times\mathbb{R}_+, \tag{3.1}$$

which embeds the PPSO problem (2.24). It is clear that we can go back to our original problem in two steps: first $v_0 = v(0, x_\alpha)$, and then $V_0 = a_0 v_0$. In the presence of management fees, by the same procedure we obtain the analogue of (3.1):

$$v(t,x) = \sup_{0\leq\tau\leq T-t} \mathsf{E}\Big[h(X^x_{\tau\wedge\tau^\dagger}) - \int_0^{\tau\wedge\tau^\dagger}\Big(p+qe^{-X^x_s}\Big)\mathrm{d}s\Big], \quad (t,x) \in [0,T]\times\mathbb{R}_+. \tag{3.2}$$

The study of (3.1) and (3.2) are equivalent from the methodological point of view (see Section 6 for a detailed discussion) and therefore we focus on (3.1) in the interest of notational simplicity. We show in Section 6 (Figure 4) how the addition of management fees affects the qualitative properties of the surrender policy.

3.1. **Theoretical results: value function and optimal exercise boundary.** In this section we provide the main theoretical results of the paper, whose proofs are given at the end of Section 5 and build upon technical results obtained in Sections 4 and 5 for the ease presentation. Let $\mathcal{L}$ be the second order differential operator associated to the diffusion (2.20), i.e.

$$(\mathcal{L}f)(x) := \tfrac{\sigma^2}{2}\partial_{xx}f(x) + \pi(x)\partial_x f(x), \quad \text{for any } f \in C^2(\mathbb{R}_+), \tag{3.3}$$

with ∂_x and ∂_{xx} denoting the first and second order partial derivatives with respect to x, respectively. We shall also denote the partial derivative with respect to time by ∂_t. As usual in optimal stopping theory, let us introduce

$$\mathcal{C} = \big\{(t,x) \in [0,T]\times\mathbb{R}_+ \ : \ v(t,x) > h(x)\big\}, \tag{3.4}$$

and

$$\mathcal{S} = \big\{(t,x) \in [0,T]\times\mathbb{R}_+ \ : \ v(t,x) = h(x)\big\}, \tag{3.5}$$

that are the so-called continuation and stopping regions, respectively. For future reference let $\partial\mathcal{C}$ be the boundary of the set $\mathcal{C}$ (notice that $\{T\}\times\mathbb{R}_+ \subseteq \partial\mathcal{C}$) and introduce the first entry time of $(t+s, X_s)$ into $\mathcal{S}$, i.e.

$$\tau^*(t,x) := \inf\big\{s \in [0, T-t] \ : \ (t+s, X^x_s) \in \mathcal{S}\big\}. \tag{3.6}$$

On the value function v of (3.1) we have the next result.

Theorem 3.1 (The value function)**.** *The function v is non-negative, continuous and bounded by 1 on the set $[0,T]\times\mathbb{R}_+$, with $v \geq h$. The mappings $t \mapsto v(t,x)$ and $x \mapsto v(t,x)$ are both non-increasing. Moreover, $v \in C^1\big([0,T)\times(0,\infty)\big)$, the second*

derivative $\partial_{xx}v$ exists and is continuous on the set $\overline{\mathcal{C}} \cap \big([0,T) \times (0,\infty)\big)$ and v solves (uniquely) the free boundary problem

$$\begin{cases} \partial_t v + \mathcal{L}v = 0, & \text{in } \mathcal{C}, \\ \partial_t v + \mathcal{L}v \leq 0, & \text{in } \mathcal{S}, \\ v \geq h, & \text{on } [0,T] \times \mathbb{R}_+, \\ v = h, & \text{on } \partial\mathcal{C}. \end{cases} \tag{3.7}$$

In the PDE literature, the above result is often presented in terms of a variational inequality. That is, v is the unique solution, in the a.e. sense of the obstacle problem:

$$\max\{\partial_t v + \mathcal{L}v, h - v\} = 0, \quad \text{on } [0,T) \times \mathbb{R}_+ \,,$$

with boundary conditions $v(T,x) = h(x)$ for $x \in \mathbb{R}_+$ and $v(t,0) = h(0)$ for $t \in [0,T)$. Uniqueness in Theorem 3.1 refers to the class of continuous functions w such that $w \in C^1([0,T) \times (0,\infty))$ with $w_{xx} \in L^\infty_{\ell oc}([0,T) \times (0,\infty))$.

From continuity of v we deduce that $\mathcal{C}$ is open and $\mathcal{S}$ is closed. Hence in particular $\partial\mathcal{C} \subset \mathcal{S}$. Moreover, standard optimal stopping results (see [30, Cor. 2.9, Sec. 2]) guarantee that the entry time τ_* to the stopping set $\mathcal{S}$ (3.6) is optimal for $v(t,x)$. It is then of interest to determine the geometry of $\mathcal{S}$.

At first we notice that since $t \mapsto v(t,x) - h(x)$ is non-increasing, then we can define

$$c(x) := \inf\{t \in [0,T] : v(t,x) = h(x)\}, \quad \text{for } x \in \mathbb{R}_+ \text{ (recall } \mathbb{R}_+ := [0,\infty)). \tag{3.8}$$

This gives us a parametrisation of the stopping set as

$$\mathcal{S} = \{(t,x) \in [0,T] \times \mathbb{R}_+ : t \geq c(x)\}. \tag{3.9}$$

In optimal stopping theory and its financial applications it is often preferable to describe the set $\mathcal{S}$ in terms of time-dependent boundaries. So, in our analysis in Sections 4–5 we use the boundary $c(\,\cdot\,)$ as a useful technical tool but we are also able to prove that it can be inverted locally and we present our results here in terms of time-dependent boundaries b_1, b_2 and b_3.

It turns out that there are two different shapes of $\mathcal{S}$ depending on the model parameters. Recalling $\bar{x}_0 = \beta + \frac{r}{\delta}$ and $x_\alpha = \ln(1/\alpha)$ we will address separately the cases $x_\alpha < \bar{x}_0$ and $x_\alpha \geq \bar{x}_0$. Our second main result is summarised below, where we denote $f(t-)$ the left limit of a function f at a point t and we adopt the convention $[t,t) = \varnothing$ for any $t \in \mathbb{R}$.

Theorem 3.2 (The optimal boundary)**.** *The following holds:*

(a) *If $x_\alpha \geq \bar{x}_0$, there exists a function $b_1 : [0,T) \to [0,\bar{x}_0]$ and a constant $t_0 \in [0,T)$ such that $b_1(t) = 0$ for $t \in [0,t_0)$, b_1 is strictly increasing and continuous on $[t_0,T)$ with $b_1(T-) = \bar{x}_0$, and the stopping region is of the form*

$$\mathcal{S} = \big\{(t,x) \in [0,T) \times \mathbb{R}_+ : x \leq b_1(t)\big\} \cup \big(\{T\} \times \mathbb{R}_+\big).$$

Thus the optimal stopping time τ_* reads

$$\tau^* = \inf\{t \in [0,T) \,:\, X_t \leq b_1(t)\} \wedge T.$$

(b) *If $x_\alpha < \bar{x}_0$, there exist constants $t_0 \in [0,T)$ and $\hat{c} \in [0,T]$, and functions $b_1 : [0,T) \to [0,x_\alpha]$ and $b_2, b_3 : [\hat{c},T) \to [x_\alpha,\bar{x}_0]$, such that:*

i) *$b_1(t) = 0$ for $t \in [0,t_0)$ and b_1 is strictly increasing and continuous on $[t_0,T)$;*

ii) *b_2 is strictly decreasing and continuous while b_3 is strictly increasing and continuous on $[\hat{c},T)$;*

iii) *$b_1(t) \leq b_2(t) \leq b_3(t)$ for $t \in [\hat{c},T)$ with $b_2(\hat{c}) = b_3(\hat{c})$ if $\hat{c} > 0$, and $\hat{c} = 0$ if $b_2(\hat{c}) < b_3(\hat{c})$;*

iv) $b_1(T-) = b_2(T-) = x_\alpha$ *and* $b_3(T-) = \bar{x}_0$*;*

v) *the stopping region is of the form*

$$\mathcal{S} = \big\{(t,x) \in [0,T) \times \mathbb{R}_+ : x \le b_1(t) \ \text{ or } x \in [b_2(t), b_3(t)]\big\} \cup \big(\{T\} \cup \mathbb{R}_+\big),$$

thus the optimal stopping time τ_* *reads*

$$\tau^* = \inf\{t \in [0,T) \,:\, X_t \le b_1(t) \ \text{ or } X_t \in [b_2(t), b_3(t)]\} \wedge T.$$

Remark 3.3. *A close inspection of the theorem above shows that* $[0,T) \times \{x_\alpha\} \subset \mathcal{C}$ *in all cases (see Proposition 5.2 for the proof).*

As anticipated in the Introduction, in case (b) we have two 'stop-loss' boundaries (i.e., b_1 and b_3), which trigger the surrender option when the BDR process crosses them downwards, and a 'too-good-to-persist' boundary (i.e., b_2), which triggers the surrender option when the BDR process crosses it upwards. In case (a) instead we only observe a single stop-loss boundary. These results will be interpreted in Section 3.3 below.

At the technical level, the strict monotonicity and continuity of the time-dependent boundaries in (a) and (b) of the theorem above are derived by analogous properties for the x-dependent boundary from (3.8). In particular, in case (a) we will prove that the function $x \mapsto c(x)$ is continuous on $(0,\infty)$, there exists $x_1 \in [0,\bar{x}_0)$ such that $c(x) = 0$ for $x \in [0,x_1)$ and it is strictly increasing on $(x_1,\bar{x}_0)$. So we have

$$b_1(t) := \inf\{x \in [0,\bar{x}_0) : c(x) > t\}, \qquad t \in [0,T).$$

In case (b) instead the geometry is more involved. We will prove that there exist $x_1 \in [0,x_\alpha)$ and $x_\alpha < x_2 \le x_3 < \bar{x}_0$ such that:

(i) $x \mapsto c(x)$ is continuous on $(0,\infty)$, $c(x) = 0$ for $x \in [0,x_1)$ and it is strictly increasing on (x_1,x_α);

(ii) the set of minimisers of $x \mapsto c(x)$ in $(x_\alpha,\bar{x}_0)$ is the closed interval $[x_2,x_3] \subset (x_\alpha,\bar{x}_0)$ where $c(\,\cdot\,)$ takes the value $\hat{c}$; if $x_2 < x_3$ then $\hat{c} = 0$;

(iii) $c(\,\cdot\,)$ is strictly decreasing on (x_α,x_2) and strictly increasing on $(x_3,\bar{x}_0)$.

Then, the boundaries b_1, b_2 and b_3 are obtained as

$$\begin{aligned} b_1(t) &:= \inf\{x \in [0,x_\alpha) : c(x) > t\}, && t \in [0,T),\\ b_2(t) &:= \sup\{x \in (x_\alpha,x_2) : c(x) > t\}, && t \in [\hat{c},T),\\ b_3(t) &:= \inf\{x \in (x_3,\bar{x}_0) : c(x) > t\}, && t \in [\hat{c},T). \end{aligned}$$

See Figures 2, 3 and 4 for various illustrations with both $\hat{c} = 0$ and $\hat{c} > 0$ and with $t_0 = 0$ and $t_0 > 0$ (notice that $t_0 = \lim_{x\downarrow 0} c(x)$).

3.2. **Some technical remarks.** The choice to work with the boundary $x \mapsto c(x)$ in our theoretical analysis is dictated by the fact that *a priori* it seems too difficult to establish existence of the three boundaries b_1, b_2 and b_3. Indeed this would normally require to prove piecewise monotonicity of the map $x \mapsto v(t,x) - h(x)$ and/or convexity of the map $x \mapsto v(t,x)$, plus developing arguments that guarantee non-emptyness of the set $\{x \in \mathbb{R}_+ : v(t,x) = h(x)\}$ depending on the choice of $t \in [0,T)$. Neither of these tasks follows by standard arguments because of the lack of an explicit solution for the SDE (2.20) and due to the absorption at $x = 0$ for the dynamics of X.

The probabilistic proof of the strict monotonicity of the optimal exercise boundaries in Theorem 3.2 is an interesting technical result in its own right and so far it was missing from the optimal stopping literature. In the PDE literature strict monotonicity of free boundaries (and even their smoothness) are well-known results. Classical references are the monographs [26] and [21] for a general treatment, while for parabolic problems with one spatial dimension (i.e., closer to our set-up) one can also refer to [7] and [20]. The

techniques developed in those seminal contributions were then employed and tailored for optimal stopping problems in mathematical finance as for example in [5, 8] for American option pricing and [11] for optimal investment with transaction costs.

For the smoothness of the boundary (understood as its continuous differentiability or higher), PDE arguments often require smoothness of the obstacle (i.e., the option's payoff) and in all cases continuous differentiability of the coefficients of the SDE underlying the stochastic optimisation. When the obstacle is not smooth (as in the American put problem) one often takes advantage of the explicit transition density of the underlying stochastic process (typically a geometric Brownian motion). In our case we have neither a smooth payoff (see (2.15)) nor continuously differentiable coefficients (see (2.21)). Moreover, the transition density of our process X is not known. So we cannot apply classical results from the PDE literature to derive continuous differentiability of the optimal boundary and we set this question aside.

For the strict monotonicity of free boundaries the PDE literature relies upon an application of Hopf's lemma and an argument by contradiction. Our probabilistic proof complements those PDE techniques by employing methods more familiar to probabilists working on optimal stopping.

The shape of the stopping region in (b) of Theorem 3.2 is somewhat remarkable and it was never observed in the context of participating policies with surrender options. Not only the stopping region is disconnected, but when $\hat{c} > 0$ there is a point in the stopping region at which one of the stop-loss boundaries meets the too-good-to-persist boundary (see Figure 1). Similar geometries of optimal stopping regions in the time-space plane have been observed numerically (see, e.g., [17, Fig.4]) but a complete theoretical analysis is not usually available. An instance of such study, revealing a similar geometry in a finite horizon optimal stopping problem, is [16]. However, the problem studied in [16] concerns the optimal prediction of the maximum of a Brownian motion with drift, whereas the one studied in [17] concerns stopping of a partially observable Brownian bridge. Hence, the similarities with the stopping rule in our set-up appear to be a mere coincidence.

Before presenting the full proofs of Theorems 3.1 and 3.2, in the next section we discuss in detail the financial interpretation of our results with the aid of extensive numerical tests. The complete theoretical analysis that leads to Theorems 3.1 and 3.2 is performed in Sections 4 and 5 for the interested reader.

3.3. **Numerical results and financial interpretation.** In order to investigate the shape of the continuation and stopping regions we implement a binomial-tree algorithm based on the diffusion approximation scheme proposed in [28]. We take a partition of $[0, T]$ with $N + 1$ equally spaced time points. At each node in the tree we associate a value of the underlying process X and of the corresponding time: that is, in the (n, j)-node we have the couple (n, x_n^j) for $j = 0, 1, \dots n$. At the subsequent time-step the process can move to one of the two nodes $(n + 1, x_n^j \pm \sigma\sqrt{\Delta})$ with $\Delta := T/N$, so that the tree is recombining. If $x_n^j > 0$ the probability p_n^j of moving upwards from the n-th node is calculated as in [28] as $p_n^j = 0 \vee [1 \wedge (\frac{1}{2} + \sqrt{\Delta} \cdot \pi(x_n^j)/2\sigma)]$. If instead $x_n^j \leq 0$ the process can only move to $(n + 1, 0)$ with probability one. We compute the numerical approximation of the value function $\tilde{v}_n(x_n^j)$ of the PPSO with the usual backward recursion, starting from $\tilde{v}_N(x_N^j) = h(x_N^j)$ for $x_N^j \geq 0$. For any $n < N$, if $x_n^j \leq 0$, then $\tilde{v}_n(x_n^j) = h(0) = 1$; if instead $x_n^j > 0$, then $\tilde{v}_n(x_n^j) = \max\{h(x_n^j), \mathsf{E}[\tilde{v}_{n+1}(X_{n+1})|X_n = x_n^j]\}$. Since the binomial-tree has recombining nodes, the evaluation of the continuation value $\mathsf{E}[\tilde{v}_{n+1}(X_{n+1})|X_n = x_n^j]$ reduces to the average of the payoff at the next two nodes.

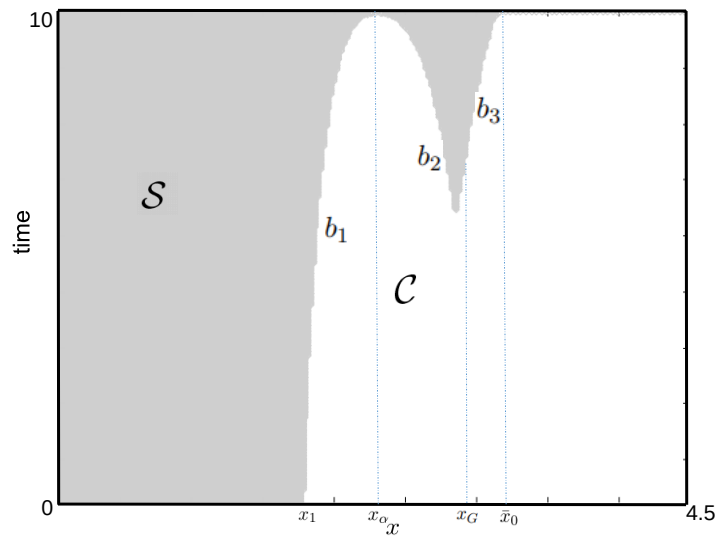

FIGURE 1. The optimal surrender regions and boundaries in the case $x_1 = b_1(0) > 0$ and $b_2(\hat{c}) = b_3(\hat{c})$ with $\hat{c} > 0$ (see Theorem 3.2-(b)).

Remark 3.4. *Notice that the regularity we have obtained for the value function v allows us, in principle, to obtain an integral equation for the optimal boundary (see [30] for some examples). However, as the explicit form of the transition density of the process X is not known, solving such integral equation numerically would not be possible. This motivates our use of binomial-trees.*

Unless otherwise specified, in the rest of the section we set the following values for the parameters (time is expressed in years while r, r^G and σ are annual rates)

$$T = 10,\ r = 1.5\%,\ \sigma = 18\%,\ r^G = 1\%,\ \delta = 0.1,\ \gamma = 0.4,\ \beta = 3,\ \alpha = 0.1. \tag{3.10}$$

For such parameter values $x_\alpha = 2.3$, $\bar{x}_0 = 3.15$ and $x_G = 3.1$ (see (2.13) and (2.22) respectively), therefore we are in the setting of $x_\alpha < \bar{x}_0$ (see (b) in Theorem 3.2).

3.3.1. *Financial interpretation of the surrender region.* We recall that the structural properties of the intrinsic value of the policy (i.e., the function h in (2.24)) and the initial value of the reserve R_0 are determined by the choice of α. From the financial perspective the PPSO is priced for each fixed value of the parameter α, that is $V_0(\alpha) = a_0 v_0(\alpha)$ (recall (2.8) and (2.24)), and the policyholder's initial BDR process at time zero is $X_0 = x_\alpha$. Figure 1 shows the optimal surrender region $\mathcal{S}$ and the boundary $c(\cdot)$ as in (3.9) on the (x, t) plane.

Optimal surrender and default. Notice that in Figure 1 the solvency requirement is always fulfilled if the policyholder exercises the SO optimally (i.e., $\tau^* < \tau^\dagger$). This corresponds to $t_0 = 0$ in (i) of Theorem 3.2-(b) and, in particular, $x_1 := b_1(0) > 0$ (compare also with x_1 in Proposition 5.3-(i)). Hence the optimal termination of the contract can only occur due to surrender or at maturity. A different situation appears in Figure 2-(i), where $t_0 = c(0+) > 0$ and early termination due to solvency requirements may occur if the dynamics of X hits zero prior to time t_0.

Minimum rate guarantee, bonus rate and a stop-loss boundary. There is a natural interpretation for the shape of the surrender region for X close to zero and for $X \geq \bar{x}_0$. On the one hand, when X is close to 0, $c(A, R) = r^G$ and $g(A, R) = R$ (see (2.4) and (2.6)

respectively); thus, the policy reserve grows at rate r^G which is lower than the discount rate r used in (2.8). So the policyholder has an incentive to surrender in order to avoid an erosion of the present value of the reserve (which is due to the gap $r^G - r < 0$). Hence it is natural to interpret the boundary b_1 as a stop-loss boundary. On the other hand, for values of X larger than $\bar{x}_0$, the policy reserve grows at a rate greater than r, due to the bonus mechanism in (2.4). In this case the policyholder has no incentive to surrender the contract and the stopping region disappears.

Stop-loss and too-good-to-persist boundaries for $X \in (x_\alpha, \bar{x}_0)$. The peculiar shape of the surrender region between x_α and $\bar{x}_0$ can be explained as follows. The value x_α is the critical value at which the participation bonus in the intrinsic value of the policy becomes active (see (2.6)). Then, if $X = x_\alpha$, the investor delays the surrender with a view to possibly receiving the bonus. Moreover, in a neighbourhood of x_α the drift in the dynamics (2.20) is positive (hence pulling the BDR process towards the bonus), so that the policyholder has an incentive to wait also if $X < x_\alpha$ but not too small. When $X > x_\alpha$ the participation bonus in the policy's intrinsic value is active and can be collected by the policyholder upon immediate surrender. This gives origin to the too-good-to-persist boundary b_2. When $X > x_G$ the bonus on the policy reserve's growth rate is also active (see (2.4)) and creates an incentive to wait. If the BDR process is close to $\bar{x}_0$ surrendering is not appealing. In fact the policyholder stays in the contract hoping that X will exceed $\bar{x}_0$ and the reserve will grow at a higher rate than the risk-free rate. However, if X decreases, the participation bonus in the intrinsic value of the policy lessens. At the same time, for $X \leq \bar{x}_0$ the growth rate of the reserve is still smaller than the risk free rate. So, as the maturity approaches, the combined effect of these two mechanisms creates the stop-loss boundary b_3.

3.3.2. *Sensitivity analysis.* Here we discuss the impact of various parameters on the shape of the surrender region and on the value of both the policy and the surrender option. In what follows both the parametrisations of the boundary $\partial\mathcal{C}$ in terms of c and in terms of b_1, b_2, b_3 will be used.

The α fraction of the initial portfolio. Figure 2 shows possible shapes of the optimal surrender boundary $c(\cdot)$ that complement the one presented in Figure 1. The plots are obtained for several values of the parameter α, or equivalently x_α (see (2.13)). The possible presence of a portion of the continuation region below the local minimum of $c(\,\cdot\,)$ (as in Figure 1) and the value $\hat{c}$ of the minimum itself depend on several factors, including the value of α. Large values of α push the initial BDR $X_0 = x_\alpha$ towards zero so that the chances of benefiting from the bonus on the policy reserve's growth rate (2.4) are slim and the policyholder will prioritise the participation bonus in the intrinsic value of the policy (2.6). That widens the area above the local minimum of $c(\,\cdot\,)$ until the continuation region (3.4) becomes completely disconnected (see Figure 2-(i)). It should be emphasised that since $X_0 = x_\alpha$, a completely disconnected surrender region means that the policyholder will surrender the contract as soon as the BDR process leaves the continuation region between the lower stop-loss boundary b_1 and the too-good-to-persist boundary b_2. As α increases this mechanism is reversed: the activation thresholds x_α of the participating bonus and x_G of the bonus on the reserve's growth rate become closer. Then the portion of stopping region above the local minimum of $c(\,\cdot\,)$ shrinks as x_α approaches $\bar{x}_0$ from the left (see Figure 2-(ii) and Figure 1 where $x_\alpha = 2.3$). In the limit we arrive to the case of $x_\alpha \geq \bar{x}_0$ (cf. (5.6)), which is also illustrated in Figure 2-(iii). There the situation is less involved because the participating bonus kicks in after the process X has already exceeded $\bar{x}_0$, so that the reserve is already growing at a rate

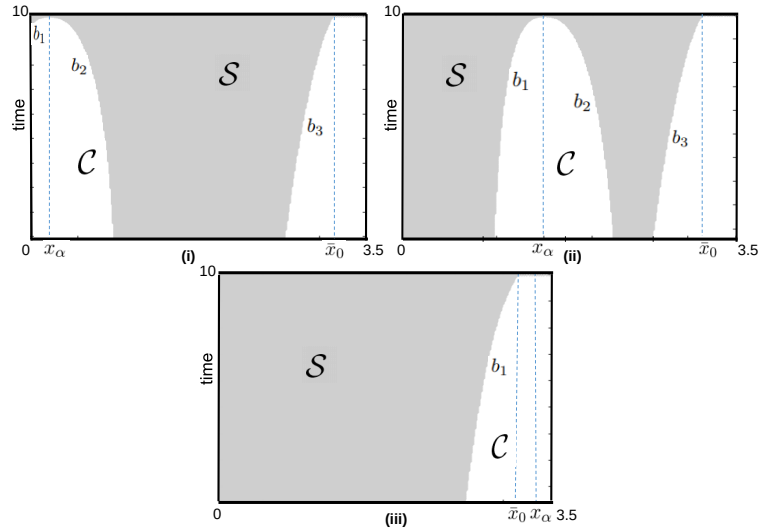

FIGURE 2. The optimal surrender boundary when varying x_α. Here $\bar{x}_0 = 3.15$ and **(i)** $x_\alpha = 0.2$, **(ii)** $x_\alpha = 1.5$, **(iii)** $x_\alpha = 3.3$.

higher than the discount rate. In that case the policyholder's waiting strategy is aimed at collecting both a large reserve and the participating bonus. The exercise of the SO in this setting is only optimal when X is sufficiently small and it is purely triggered by the stop-loss mechanism due to discounting.

Participation coefficient and minimum rate guarantee. In Figure 3 we study the sensitivity of the optimal surrender boundary $c(\cdot)$ with respect to γ (left plot) and to r^G (right plot). We remark that γ only affects the intrinsic value of the policy (see (2.6)) but it does not affect either the reference portfolio A nor the policy reserve R. As γ increases, the participating bonus increases and counters the effect of discounting. So the policyholder is inclined to stay in the contract longer to see if the the bonus mechanism on the reserve will also be activated. Likewise, as r^G increases the policyholder has progressively more benefits from staying in the contract, then the SO becomes less appealing and the area between the boundaries b_2 and b_3 shrinks. In particular, as r^G approaches the risk free rate r the stop-loss boundary b_3 and the too-good-to-persist boundary b_2 do not disappear but become less extended (theoretically this is expected because the interval $(x_\alpha, \bar{x}_0)$ does not depend on r^G and it is shown in Lemma 5.4 that $\mathcal{S} \cap \big([0,T) \times (x_\alpha, \bar{x}_0)\big) \neq \varnothing$). In both situations the incentive to surrender the contract decreases and the continuation region expands. As a result the optimal boundary $c(\cdot)$ is pushed upwards in our plots.

Values of the policy and of the surrender option. We conclude the section by analysing how the bonus distribution mechanism and the minimum interest rate guarantee in the policy reserve impact on the value of the policy and on the value of the embedded SO. The value of the SO is obtained, as in (2.9), by comparing the value of the PPSO to the value of its European counterpart (i.e., with no SO). Now we fix the initial portfolio value $A_0 = 1,000$ so that $R_0 = 100$. We collect in Table 1 the value V_0 of the PPSO (see (2.8)), the value V_0^E of the contract without SO (see (2.10)) and the value V^{opt} of the SO. As in [22], we consider the following three scenarios depending on the level of participation in the returns generated by the reference portfolio: *low* ($\delta = 0.1$ and $\beta = 3.4$), *medium* ($\delta = 0.25$ and $\beta = 2.7$), *high* ($\delta = 0.6$ and $\beta = 2$). The lower the

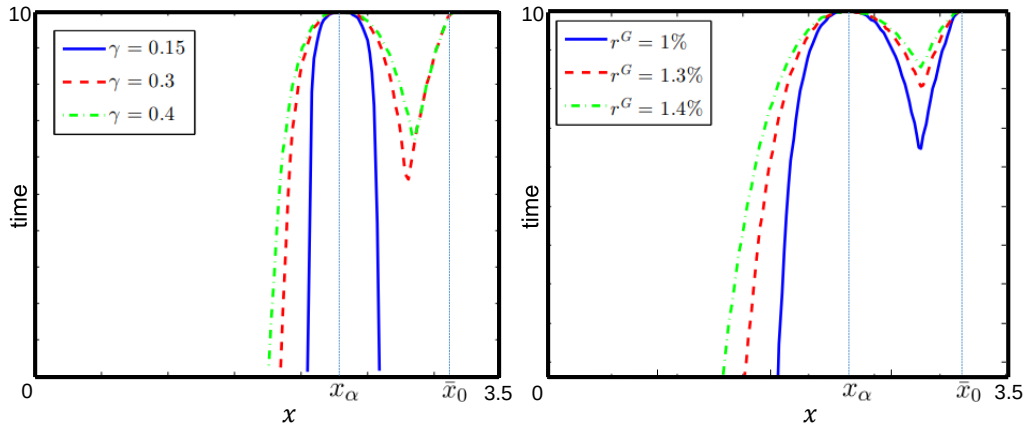

FIGURE 3. Sensitivity of the optimal surrender boundary $c(\cdot)$ with respect to γ (left plot) and to r^G (right plot). For $\gamma = 0.15$ (left plot) we omit the rightmost portion of the boundary which is never reached when $X_0 = x_\alpha$ and the policyholder stops optimally.

value of δ, the less the policyholder participates in the reserve via (2.4). Moreover, the higher the value of the target buffer ratio β, the smaller is the surplus that the policyholder receives. The value of V_0^E is evaluated by using the same binomial-tree method described above, without the complication of the optimisation which is required at each node in the PPSO. As expected, V_0 is always greater than V_0^E. Their difference gives the option value V^{opt}.

Spread	Scenario	V_0	V_0^E	V^{opt}
$r - r^G = 0.5\%$	low	100.7	99.44	1.26
	medium	104.16	103.43	0.73
	high	160.93	160.41	0.52
$r - r^G = 0.8\%$	low	100.27	94.92	5.35
	medium	102.17	99.29	2.88
	high	158.14	156.47	1.67
$r - r^G = 1.5\%$	low	100.14	88.98	11.16
	medium	100.64	93.93	6.71
	high	154.81	151.38	3.43

TABLE 1. The values V_0 of the PPSO, V_0^E of the contract without surrender option and V^{opt} of the surrender option.

In the *low* scenario the value of the European contract V_0^E is below par for all values of the spread $r - r^G$, i.e. $V_0^E < R_0$. If the spread is relatively large, i.e. 0.8% or 1.5%, the European contract trades below par also in the *medium* scenario. When the minimum interest rate guaranteed r^G is much smaller than the market risk-free rate and the level of participation in the returns is also relatively small, the policy is not financially appealing to an investor if compared, for example, to bond investments.

However, since $R_0 > V_0^E$ (see Table 1) an investor who purchases the European contract incurs an initial outlay which is smaller than the initial amount credited to the reserve. This makes the policy potentially appealing as a form of secure savings.

The value V_0 of the PPSO is always at or above par, i.e. $V_0 \geq R_0$, due to the American-type option embedded in the contract (at par the SO is immediately exercised). Contract values V_0 and V_0^E increase moving from *low* towards *high* scenario whereas the value of V^{opt} decreases. This shows that the incentive to exercise the SO is reduced by higher participation of the investor in the returns. On the contrary, as $r - r^G$ increases the contract values decrease, whereas V^{opt} increases. This is in line with the intuition that the higher the spread, the less the contract is profitable for the policyholder, hence creating a big incentive to exercise the SO.

4. Properties of the value function

In this section we collect some facts about the underlying stochastic process X, defined in (2.20), which will then be used to infer regularity of the value function (3.1).

4.1. **Path properties of the underlying process.** First we observe that since the drift function $\pi(\,\cdot\,)$ is Lipschitz continuous and the diffusion coefficient is constant, there exists a modification $\widetilde{X}$ of X such that the stochastic flow $(t,x) \mapsto \widetilde{X}_t^x(\omega)$ is continuous for a.e. $\omega \in \Omega$ (see, e.g., [31, Chapter V.7]). As usual, throughout the paper we work with the continuous modification which we still denote by X for simplicity.

Lemma 4.1. *For any* P*-a.s. finite stopping time* $\tau \geq 0$ *it holds*

$$\big|X_\tau^x - X_\tau^y\big| \leq \big|x-y\big| e^{\delta\tau}, \quad \mathsf{P}\text{-a.s. for } x,y \in \mathbb{R}_+, \tag{4.1}$$

$$X_\tau^y - X_\tau^x \geq (y-x)(2-e^{\delta\tau}), \quad \mathsf{P}\text{-a.s. for } y \geq x \geq 0. \tag{4.2}$$

Proof. From the integral form of (2.20) (with $X_0^x = x$ and $X_0^y = y$), and noticing that $\pi(\cdot)$ is Lipschitz with constant $\delta > 0$, it is immediate to see

$$\big|X_\tau^x - X_\tau^y\big| \leq \big|x-y\big| + \int_0^\tau \big|\pi(X_t^x) - \pi(X_t^y)\big|\mathrm{d}t \leq \big|x-y\big| + \delta \int_0^\tau \big|X_t^x - X_t^y\big|\mathrm{d}t.$$

Then, an application of Gronwall's inequality gives (4.1)

The argument for (4.2) is similar. Using that $y > x$ and $\pi(\cdot)$ Lipschitz we have

$$\begin{aligned} X_\tau^y - X_\tau^x &\geq y - x - \delta \int_0^\tau \big|X_t^y - X_t^x\big|\mathrm{d}t \\ &\geq y - x - (y-x)\int_0^\tau \delta e^{\delta t}\mathrm{d}t = (y-x)(2-e^{\delta\tau}), \end{aligned}$$

where the second inequality uses (4.1). □

The next estimate on the local time of the process X is particularly useful to establish that the value function is Lipschitz in the time variable. In the rest of the paper we denote $L^z := (L_t^z)_{t\in[0,T]}$ the local time of the process X at a point $z \geq 0$, which is defined as (see, e.g., [30, Eq. (3.3.29), p. 68])

$$L_t^z(X) := \lim_{\varepsilon \downarrow 0} \frac{1}{2\varepsilon} \int_0^t \mathbb{1}_{\{|X_s - z| \leq \varepsilon\}} \mathrm{d}\langle X\rangle_s, \quad \mathsf{P}\text{-a.s.} \tag{4.3}$$

Recall that $\mathsf{E}_x[\,\cdot\,] = \mathsf{E}[\,\cdot\,|X_0 = x]$.

Lemma 4.2. *Let $0 < t_1 \leq t_2 \leq T$, fix $N > 0$ and recall x_α from (2.13). Then, there exists a positive constant $\kappa := \kappa(t_1, N; x_\alpha)$ such that*

$$\sup_{x\in[0,N]} \mathsf{E}_x \left[L^{x_\alpha}_{t_2} - L^{x_\alpha}_{t_1}\right] \leq \kappa(t_2 - t_1). \tag{4.4}$$

Proof. Thanks to (4.3) we can select a sequence $(\varepsilon_n)_{n\geq 1}$ such that $\varepsilon_n \downarrow 0$ as $n \to \infty$ and

$$L^{x_\alpha}_{t_2} - L^{x_\alpha}_{t_1} = \lim_{n\to\infty} \frac{1}{2\varepsilon_n} \int_{t_1}^{t_2} \mathbb{1}_{\{|X_s - x_\alpha| \leq \varepsilon_n\}} \mathrm{d}\langle X\rangle_s, \qquad \mathsf{P}_x - a.s.$$

Then, using Fatou's lemma we get

$$\mathsf{E}_x \left[L^{x_\alpha}_{t_2} - L^{x_\alpha}_{t_1}\right] \leq \liminf_{n\to\infty} \frac{1}{2\varepsilon_n} \int_{t_1}^{t_2} \mathsf{P}_x \left(|X_s - x_\alpha| \leq \varepsilon_n\right) \sigma^2 \mathrm{d}s. \tag{4.5}$$

It is well-known that X admits a transition density with respect to its speed measure (see, e.g., [32, Thm. 50.11] or [23, Sec. 4.11]). That is

$$\mathsf{P}_x \left(|X_s - x_\alpha| \leq \varepsilon_n\right) = \int_{x_\alpha - \varepsilon_n}^{x_\alpha + \varepsilon_n} p(s, x, y) \frac{2\mathrm{d}y}{\sigma^2 S'(y)},$$

where S' is the derivative of the scale function and reads

$$S'(y) = \exp\left(-\frac{2}{\sigma^2} \int_0^y \pi(z)\mathrm{d}z\right). \tag{4.6}$$

Moreover, the map $(s, x, y) \mapsto p(s, x, y)$ is continuous on $(0, \infty) \times \mathbb{R}^2$ and clearly S' is continuous too. Hence, letting $\varepsilon_n \leq \varepsilon_0$, for all $n \geq 1$ and some $\varepsilon_0 > 0$, and setting

$$\kappa(t_1, N; x_\alpha) := 2 \sup_{(s,x,y)} \frac{p(s, x, y)}{S'(y)},$$

with the supremum taken over $(s, x, y) \in [t_1, T] \times [0, N] \times [x_\alpha - \varepsilon_0, x_\alpha + \varepsilon_0]$, it is immediate to obtain (4.4) from (4.5). □

Remark 4.3. *Notice that in the lemma above t_1 must be taken strictly positive as the constant $\kappa(t_1, N; x_\alpha)$ might (and will) explode as $t_1 \to 0$.*

4.2. **Continuity and monotonicity of the value function.** Some parts of the analysis in our paper are more conveniently performed by considering a different formulation of problem (3.1). Recall the infinitesimal generator $\mathcal{L}$ of X (see (3.3)) and define the function

$$H(x) := \begin{cases} e^{-x}\left(\frac{\sigma^2}{2} - \pi(x)\right), & x \leq x_\alpha \\ (1-\gamma)\, e^{-x}\left(\frac{\sigma^2}{2} - \pi(x)\right), & x > x_\alpha, \end{cases} \tag{4.7}$$

with x_α from (2.13). For future reference it is worth noticing that, since $r^G < r$,

$$-(r - r^G) \leq H(x) \leq \delta \quad \text{for } x \in \mathbb{R}_+. \tag{4.8}$$

Clearly H is discontinuous at x_α and it is easy to check that $H(x) = (\mathcal{L}h)(x)$ for $x \neq x_\alpha$ (recall that h depends on α). Since $x \mapsto h(x)$ (see (2.15)) is a convex function and its first derivative has a single jump

$$\partial_x h(x_\alpha+) - \partial_x h(x_\alpha-) = \gamma\alpha,$$

we can apply Itô-Tanaka's formula to $h(X_{\tau\wedge\tau^\dagger})$ in (3.1), to obtain the following equivalent formulation of problem (3.1)

$$u(t,x) := v(t,x) - h(x) = \sup_{0\le\tau\le T-t} \mathsf{E}_x\Big[\int_0^{\tau\wedge\tau^\dagger} H(X_s)\mathbb{1}_{\{X_s\neq x_\alpha\}}\mathrm{d}s + \frac{\gamma\alpha}{2}L^{x_\alpha}_{\tau\wedge\tau^\dagger}\Big], \tag{4.9}$$

where $(L^z_t)_{t\ge0}$ is the local time of X at a point $z>0$ (see (4.3)). Notice that u is non-negative since $v(t,x)\ge h(x)$, for all $(t,x)\in[0,T]\times\mathbb{R}_+$, by (3.1). We will show in Proposition 5.2 that the presence of local time L^{x_α} in (4.9) implies that it is never optimal to stop when the process X is equal to x_α.

Proposition 4.4. *The following properties hold for the value function of the optimal stopping problem (3.1):*

i) *the map $t\mapsto v(t,x)$ is decreasing and $v(T,x)=h(x)$ for any fixed $x\ge0$;*
ii) *the map $x\mapsto v(t,x)$ is decreasing and $v(t,0)=h(0)=1$ for any fixed $t\in[0,T]$.*

Moreover, for any $0\le x_1\le x_2<+\infty$ and any $t\in[0,T]$ it holds

$$0\le v(t,x_1)-v(t,x_2)\le \kappa_0(x_2-x_1), \tag{4.10}$$

with $\kappa_0 := e^{\delta T}$. The map $t\mapsto v(t,x)$ is continuous on $[0,T]$ for any $x\in\mathbb{R}_+$ and, finally, for any $0\le t_1\le t_2<T$ and any $x\in[0,N]$, with fixed $N>0$, there is a constant $\kappa_1=\kappa_1(t_2,N;x_\alpha)>0$ such that

$$0\le v(t_1,x)-v(t_2,x)\le \kappa_1(t_2-t_1). \tag{4.11}$$

Proof. The monotonicity in point $i)$ follows from time-independence of h and $\tau^\dagger$, whereas the value of v at T follows from (3.1). As for $ii)$, $v(t,0)=h(0)$ since $\tau^\dagger(0)=0$ P-a.s. To show monotonicity of v in x, fix $x_1<x_2$ and note that by uniqueness of the solution to (2.20) follows $X^{x_1}_{s\wedge\tau^\dagger(x_1)}\le X^{x_2}_{s\wedge\tau^\dagger(x_2)}$ P-a.s. for all $s\in[0,T]$. Since the inequality also holds if we replace s by a stopping time and the gain function h is decreasing, we obtain

$$v(t,x_2)-v(t,x_1)\le \sup_{0\le\tau\le T-t}\mathsf{E}\Big[h(X^{x_2}_{\tau\wedge\tau^\dagger(x_2)})-h(X^{x_1}_{\tau\wedge\tau^\dagger(x_1)})\Big]\le 0. \tag{4.12}$$

Next we prove (4.10). Fix $t\in[0,T]$, consider $0\le x_1<x_2$ and denote by $\tau_1^\dagger := \tau^\dagger(x_1)$ and $\tau_2^\dagger := \tau^\dagger(x_2)$ the first hitting time at zero of X^{x_1} and X^{x_2}, respectively. From pathwise uniqueness of the solution of (2.20) we have $\tau_1^\dagger\le\tau_2^\dagger$. Then $\tau\wedge\tau_1^\dagger\wedge\tau_2^\dagger=\tau\wedge\tau_1^\dagger$, P-a.s. for every admissible stopping time τ. Recalling that $v(t,\cdot)$ is decreasing and that h is strictly decreasing and 1-Lipschitz (see (2.16)) we have

$$\begin{aligned} 0\le v(t,x_1)-v(t,x_2) &\le \sup_{0\le\tau\le T-t}\mathsf{E}\Big[h\big(X^{x_1}_{\tau\wedge\tau_1^\dagger}\big)-h\big(X^{x_2}_{\tau\wedge\tau_1^\dagger}\big)\Big] \\ &\le \mathsf{E}\Big[\sup_{0\le s\le T-t}\Big|X^{x_1}_s-X^{x_2}_s\Big|\Big]\le e^{\delta T}(x_2-x_1), \end{aligned} \tag{4.13}$$

where the first inequality is obtained by taking $\tau\wedge\tau_1^\dagger$ in $v(t,x_2)$ and the last inequality follows by (4.1).

It remains to prove (4.11). For that it is convenient to use (4.9) and notice that for $0\le t_1\le t_2<T$ and $x\in\mathbb{R}_+$ we have

$$0\le v(t_1,x)-v(t_2,x)=u(t_1,x)-u(t_2,x),$$

where the inequality is due to $i)$ above. For any stopping time $\tau\in[0,T-t_1]$ we have that $\tau\wedge(T-t_2)$ is admissible for the problem with value $u(t_2,x)$. Then, by direct

comparison (recall that $\tau^\dagger$ only depends on $x \in \mathbb{R}_+$) and with $x \in [0, N]$, we have

$$\begin{aligned} 0 \leq & u(t_1, x) - u(t_2, x) \\ \leq & \sup_{0 \leq \tau \leq T - t_1} \mathsf{E}_x \left[\mathbb{1}_{\{\tau \wedge \tau^\dagger > T - t_2\}} \left(\int_{T-t_2}^{\tau \wedge \tau^\dagger} \mathbb{1}_{\{X_s \neq x_\alpha\}} H(X_s) \mathrm{d}s + \frac{\alpha\gamma}{2} \left(L^{x_\alpha}_{\tau \wedge \tau^\dagger} - L^{x_\alpha}_{T-t_2} \right) \right) \right] \\ \leq & \, \delta(t_2 - t_1) + \frac{\alpha\gamma}{2} \mathsf{E}_x \left[L^{x_\alpha}_{T-t_1} - L^{x_\alpha}_{T-t_2} \right], \end{aligned} \tag{4.14}$$

where in the final inequality we used (4.8) and the fact that the local time $t \mapsto L^{x_\alpha}_t$ is non-decreasing. Continuity of $t \mapsto u(t, x)$ (hence of $t \mapsto v(t, x)$) is now clear by continuity of local time's sample paths. Recalling Lemma 4.2 we also obtain (4.11) by setting $\kappa_1(t_2, N; x_\alpha) := \delta + \alpha\gamma/2 \cdot \kappa(T - t_2, N; x_\alpha)$. □

An immediate consequence of the proposition above, and the fact that h is bounded and non-negative, is given in the next corollary.

Corollary 4.5. *The value function v of the optimal stopping problem (3.1) is non-negative, continuous on $[0, T] \times \mathbb{R}_+$ and bounded by 1.*

Recalling the sets $\mathcal{C}$ and $\mathcal{S}$ defined in (3.4) and (3.5), we can express them in terms of the function u from (4.9) as

$$\mathcal{C} = \left\{ (t, x) \in [0, T] \times \mathbb{R}_+ \ : \ u(t, x) > 0 \right\}$$

and

$$\mathcal{S} = \left\{ (t, x) \in [0, T] \times \mathbb{R}_+ \ : \ u(t, x) = 0 \right\}.$$

Continuity of v and h imply that the sets $\mathcal{C}$ and $\mathcal{S}$ are open and closed, respectively. Moreover, [30, Cor. 2.9, Sec. 2] guarantees that τ_* defined in (3.6) is optimal for $v(t, x)$ for any $(t, x) \in [0, T] \times \mathbb{R}_+$. Finally, [30, Thm. 2.4, Sec. 2] ensures that the process $V^{t,x} := (V^{t,x}_s)_{s \in [0, T-t]}$ given by $V^{t,x}_s = v(t+s, X^x_s)$ is a supermartingale while $(V^{t,x}_{s \wedge \tau^*})_{s \in [0, T-t]}$ is a martingale, for any $(t, x) \in [0, T] \times \mathbb{R}_+$. Using the martingale property and continuity of the value function we obtain the next well-known result (see, e.g. [30, Sec. 7.1, Chapter III], for a proof).

Proposition 4.6. *The value function v lies in $C^{1,2}(\mathcal{C})$ and it solves the boundary value problem*

$$\partial_t v + \mathcal{L} v = 0, \qquad \text{in } \mathcal{C} \tag{4.15}$$

with $v = h$ on $\partial\mathcal{C}$.

The next simple technical lemma is consequence of the maximum principle and it will be used later to prove continuity and strict monotonicity of the stopping boundary.

Lemma 4.7. *For all $(t, x) \in \mathcal{C}$ it holds $\partial_t v(t, x) < 0$.*

Proof. By contradiction we assume there is $(t_0, x_0) \in \mathcal{C}$ such that $\partial_t v(t_0, x_0) = 0$. Since $v(t_0, x_0) > h(x_0)$ and $v(T, x_0) = h(x_0)$, by continuity there must exists $t_1 \in (t_0, T)$ such that $(t_1, x_0) \in \mathcal{C}$ and $\partial_t v(t_1, x_0) < -\varepsilon$, for some $\varepsilon > 0$. By continuity of $\partial_t v$ inside $\mathcal{C}$, and the fact that $\mathcal{C}$ is open, there exists $\delta > 0$ such that $\partial_t v(t_1, x) < -\varepsilon/2$ for $x \in (x_0 - \delta, x_0 + \delta)$ and $\{t_1\} \times (x_0 - \delta, x_0 + \delta) \subset \mathcal{C}$.

Now, letting $\mathcal{O} := (t_0, t_1) \times (x_0 - \delta, x_0 + \delta)$ we have that $\mathcal{O} \subseteq \mathcal{C}$ and $\partial_t v \in C^{1,2}(\mathcal{O})$, thanks to internal regularity results for solutions of partial differential equations applied to (4.15) (see, e.g., [19, Thm. 10, Ch. 3, Sec. 5]). Moreover, differentiating (4.15) with

respect to time and using Proposition 4.4-(i) with the observations above, we obtain that $\hat{v} := \partial_t v$ solves

$$(\partial_t \hat{v} + \mathcal{L}\hat{v})(t,x) = 0, \qquad \text{for } (t,x) \in \mathcal{O} \tag{4.16}$$

$$\hat{v}(t, x_0 \pm \delta) \leq 0, \qquad \text{for } t \in [t_0, t_1) \tag{4.17}$$

$$\hat{v}(t_1, x) < -\varepsilon/2, \qquad \text{for } x \in (x_0 - \delta, x_0 + \delta). \tag{4.18}$$

Setting $\tau_{\mathcal{O}} := \inf\{s \geq 0 \,:\, (t_0 + s, X_s^{x_0}) \notin \mathcal{O}\}$, an application of Dynkin's formula gives

$$0 = \hat{v}(t_0, x_0) = \mathsf{E}\left[\hat{v}(t_0 + \tau_{\mathcal{O}}, X_{\tau_{\mathcal{O}}}^{x_0})\right] \leq -\frac{\varepsilon}{2}\mathsf{P}\big(\tau_{\mathcal{O}} = t_1 - t_0\big), \tag{4.19}$$

which leads to a contradiction as the process $(t_0 + s, X_s^{x_0})$ exits $\mathcal{O}$ by crossing the segment $\{t_1\} \times (x_0 - \delta, x_0 + \delta)$ with positive probability. □

It is clear by (4.9) that u inherits the same continuity and boundedness properties of v (see (4.11), (4.10) and Corollary 4.5). Moreover, $\partial_t u < 0$ in $\mathcal{C}$ with $u \in C^{1,2}$ in $\mathcal{C} \setminus ([0,T] \times \{x_\alpha\})$ due to (2.15) and (4.7). Finally, in $\mathcal{C} \setminus ([0,T] \times \{x_\alpha\})$ the function u solves

$$\partial_t u + \mathcal{L}u = -H, \tag{4.20}$$

with $u = 0$ on $\partial\mathcal{C}$.

5. The free boundary problem

In this section we study the free boundary problem associated with the stopping problem (4.9). We derive geometric properties of the continuation region $\mathcal{C}$ and regularity of its boundary $\partial\mathcal{C}$. These have a close interplay with the smoothness of the value function v in the whole space.

5.1. Analysis of the stopping region.

We start the study of the stopping region by noting that for any $(t,x) \in [0,T] \times \mathbb{R}_+$ it holds

$$(t,x) \in \mathcal{S} \Rightarrow [t,T] \times \{x\} \in \mathcal{S}, \tag{5.1}$$

since $t \mapsto u(t,x)$ is non-increasing (see (i) in Proposition 4.4).

Some of the arguments that we need in order to characterise the stopping region require the next lemma. Its proof is somewhat standard but we provide it in the Appendix for completeness.

Lemma 5.1. *For $\varepsilon > 0$ define*

$$\rho_\varepsilon := \inf\{s \geq 0 \,:\, X_s^{x_\alpha} \notin (x_\alpha - \varepsilon, x_\alpha + \varepsilon)\}.$$

Then, for any $\ell > 0$ there exists $t_{\varepsilon,\ell} > 0$ such that

$$\mathsf{E}_{x_\alpha}\left[L_{s\wedge\rho_\varepsilon}^{x_\alpha}\right] > \ell\, \mathsf{E}_{x_\alpha}\left[s \wedge \rho_\varepsilon\right] \quad \textit{for all } s \in (0, t_{\varepsilon,\ell}).$$

Now we can use the lemma to show that it is never optimal to stop at x_α.

Proposition 5.2. *It holds $[0,T) \times \{x_\alpha\} \subset \mathcal{C}$.*

Proof. Fix $\varepsilon > 0$ and let ρ_ε be as in Lemma 5.1. Take $t \in [0,T)$ and $s \in [0, T-t)$. Since stopping at $s \wedge \rho_\varepsilon$ is admissible for the problem with value function $u(t, x_\alpha)$, and

$$\inf_{|\zeta| \leq \varepsilon} H(x_\alpha + \zeta) \geq -c_\varepsilon,$$

for some $c_\varepsilon > 0$ only depending on ε, one obtains

$$\begin{aligned} u(t,x_\alpha) &\geq \mathsf{E}_{x_\alpha}\left[\int_0^{s\wedge\rho_\varepsilon} H(X_u)\mathbb{1}_{\{X_u\neq x_\alpha\}}du + \frac{\gamma\alpha}{2}L^{x_\alpha}_{s\wedge\rho_\varepsilon}\right] \\ &\geq \tfrac{\gamma\alpha}{2}\mathsf{E}_{x_\alpha}\left[L^{x_\alpha}_{s\wedge\rho_\varepsilon}\right] - c_\varepsilon\mathsf{E}_{x_\alpha}\left[s\wedge\rho_\varepsilon\right]. \end{aligned}$$

Now, applying Lemma 5.1 with $\ell = 2c_\varepsilon/(\gamma\alpha)$ and picking $s > 0$ sufficiently small gives $u(t,x_\alpha) > 0$. Hence $(t,x_\alpha) \in \mathcal{C}$. Since $t \in [0,T)$ was arbitrary, the claim follows. □

For any initial point (t,x) with $t \in [0,T)$ and $x \in \mathbb{R}_+ \setminus \{x_\alpha\}$ such that $H(x) > 0$, we can choose to stop at the first exit time from a small interval centered at x. Since $H > 0$ in such interval, and the stopping time is strictly positive P-a.s., this well-known argument gives $u(t,x) > 0$. Then, it follows that $\mathcal{R} \subseteq \mathcal{C}$, where

$$\mathcal{R} := \{(t,x) \in [0,T) \times (\mathbb{R}_+\backslash\{x_\alpha\}) : H(x) > 0\}. \tag{5.2}$$

Combining this observation with Proposition (5.2) we get

$$\mathcal{R} \cup ([0,T) \times \{x_\alpha\}) \subseteq \mathcal{C}. \tag{5.3}$$

It is clear that the shape of the set $\mathcal{R}$ varies depending on the parameters of the problem. Interestingly, this gives rise to two possible shapes of the stopping region, as we will see in the rest of the section. Let us start by noticing that

$$H(x) > 0 \iff \frac{\sigma^2}{2} - \pi(x) > 0 \iff x > \bar{x}_0 = \beta + \frac{r}{\delta}, \tag{5.4}$$

where we used (2.21) and $r > r_G$. Then, based on the fact that

$$\mathcal{S} \subseteq \big(\mathcal{R}^c \cap \{x \neq x_\alpha\}\big) \cup \big(\{T\} \times \mathbb{R}_+\big),$$

where $\mathcal{R}^c$ is the complement of $\mathcal{R}$, we distinguish two cases:

Case 1: $x_\alpha < \bar{x}_0$, then we have

$$\mathcal{S} \subseteq \Big\{[0,T) \times \Big([0,x_\alpha) \cup (x_\alpha,\bar{x}_0]\Big)\Big\} \cup \Big(\{T\} \times \mathbb{R}_+\Big). \tag{5.5}$$

Case 2: $x_\alpha \geq \bar{x}_0$, then we have

$$\mathcal{S} \subseteq \Big\{[0,T) \times [0,\bar{x}_0)\Big\} \cup \Big(\{T\} \times \mathbb{R}_+\Big). \tag{5.6}$$

We now focus on the study of the optimal stopping region in Case 1. Case 2 is easier and can be handled with simpler methods. Thanks to (5.1) we may write (recall (3.8))

$$c(x) = \inf\{t \in [0,T] \,:\, u(t,x) = 0\}, \quad \text{for } x \in \mathbb{R}_+. \tag{5.7}$$

Notice that

$$c(x_\alpha) = T \qquad \text{and} \qquad c(x) = T, \quad \text{for } x > \bar{x}_0, \tag{5.8}$$

due to (5.3). The next proposition is the main result in this subsection. It provides piecewise monotonicity and right/left-continuity of $c(\,\cdot\,)$.

Proposition 5.3. *Assume $x_\alpha < \bar{x}_0$. The map $x \mapsto c(x)$ attains a global minimum $0 \leq \hat{c} \leq T$ on $[x_\alpha,\bar{x}_0]$. Moreover, there exist $x_1 \in [0,x_\alpha)$, $x_2 \in (x_\alpha,\bar{x}_0)$ and $x_3 \in [x_2,\bar{x}_0)$ such that $c(\cdot)$ is*

(i) *equal to zero on $[0,x_1]$, strictly increasing on $(x_1,x_\alpha]$ and left-continuous on $[0,x_\alpha)$;*
(ii) *strictly decreasing on $[x_\alpha,x_2)$ and right-continuous on $[x_\alpha,x_3)$;*
(iii) *strictly increasing on $[x_3,\bar{x}_0)$ and left-continuous on $[x_2,\bar{x}_0)$.*

(Notice that in (ii) and (iii) it might be $x_2 = x_3$.)

In all cases $\hat{c} = c(x_2) = c(x_3)$ and, if $x_2 < x_3$, then $\hat{c} = 0$. Finally,

$$\lim_{x\to x_\alpha} c(x) = c(x_\alpha) = T \quad \text{and} \quad T = \lim_{x\downarrow \bar{x}_0} c(x) \geq c(\bar{x}_0) = \lim_{x\uparrow \bar{x}_0} c(x). \tag{5.9}$$

The proof relies on two technical lemmas which we are going to present first.

Lemma 5.4. *Assume $x_\alpha < \bar{x}_0$. Then*

(i) *for $z < x_\alpha$ and $t \in (0,T]$ it holds*

$$(t,z) \in \mathcal{S} \implies [t,T] \times [0,z] \subseteq \mathcal{S};$$

(ii) *for $z_1, z_2 \in (x_\alpha, \bar{x}_0)$, with $z_1 < z_2$, and $t \in (0,T]$ it holds*

$$(t,z_1),(t,z_2) \in \mathcal{S} \implies [t,T] \times [z_1,z_2] \subseteq \mathcal{S}.$$

The map $x \mapsto c(x)$ is never strictly positive and constant (simultaneously) on intervals (z_1,z_2) contained in $[0,x_\alpha) \cup (x_\alpha, \bar{x}_0)$. Finally, for every interval (z_1,z_2) contained in $[0,x_\alpha) \cup (x_\alpha,\bar{x}_0)$ it holds

$$\mathcal{S} \cap \big((0,T) \times (z_1,z_2)\big) \neq \varnothing. \tag{5.10}$$

Proof. First we prove (i) and (ii). The two claims are similar since $(t,0) \in \mathcal{S}$ for all $t \in [0,T]$ (see Proposition 4.4-(ii)). Then it is enough to show (ii) as the proof of point (i) is analogous up to obvious changes.

Let (t,z_1) and (t,z_2) belong to $\mathcal{S}$ and $x_\alpha < z_1 < z_2 \leq \bar{x}_0$. If $t = T$ the result is trivial due to (5.8). Then let $t < T$ and recall that $H(x) < 0$ for $x \in (x_\alpha, \bar{x}_0)$. By (5.1) we know that $[t,T] \times \{z_i\} \subseteq \mathcal{S}$ for $i = 1,2$. Then it suffices to show that also $\{t\} \times (z_1,z_2) \subseteq \mathcal{S}$. Arguing by contradiction assume there exists $z_3 \in (z_1,z_2)$ such that $(t,z_3) \in \mathcal{C}$. Let $\tau_3^* = \tau^*(t,z_3)$ be optimal for the problem with value $u(t,z_3)$. Then

$$u(t,z_3) = \mathsf{E}_{z_3}\bigg[\int_0^{\tau_3^*\wedge\tau^\dagger} H(X_s)\mathbb{1}_{\{X_s\neq x_\alpha\}}\mathrm{d}s + \frac{\gamma\alpha}{2}L^{x_\alpha}_{\tau_3^*\wedge\tau^\dagger}\bigg].$$

Since $[t,T] \times \{z_i\} \subseteq \mathcal{S}$ for $i = 1,2$, we have that

$$\tau_3^* \leq \zeta := \inf\{s \geq 0 \,:\, (t+s, X_s^{z_3}) \notin [t,T) \times (z_1,z_2)\}.$$

Hence, $u(t,z_3) < 0$ because for all $s \geq 0$ we have $L^{x_\alpha}_{s\wedge\zeta} = 0$ and $H(X_{s\wedge\zeta}) < 0$, P_{z_3}-a.s. and $\mathsf{P}_{z_3}(\tau_3^* > 0) = 1$ by assumption. Thus we have a contradiction.

Next we show that c cannot be strictly positive and constant. The proof borrows some ideas from [12]. Arguing by contradiction, assume that there exists an interval $(z_1,z_2) \subset [0,x_\alpha) \cup (x_\alpha,\bar{x}_0)$ where $c(x)$ takes the constant value $\bar{c} > 0$. Then the open set $\mathcal{O} := (0,\bar{c}) \times (z_1,z_2)$ is contained in $\mathcal{C}$ and $u \in C^{1,2}(\mathcal{O})$ since so are v, by Proposition 4.6, and h (away from x_α). It follows that u satisfies

$$\begin{cases} \partial_t u + \mathcal{L}u = -H, & \text{in } \mathcal{O}, \\ u(\bar{c},x) = 0, & x \in (z_1,z_2)\ . \end{cases} \tag{5.11}$$

Pick $\varphi \in C_c^\infty(z_1,z_2)$, with $\varphi \geq 0$. Thanks to (5.11), for $s \in [0,\bar{c})$ we have

$$\begin{aligned}\int_{z_1}^{z_2} \partial_t u(s,y)\varphi(y)\mathrm{d}y &= -\int_{z_1}^{z_2} (\mathcal{L}u)(s,y)\varphi(y)\mathrm{d}y - \int_{z_1}^{z_2} H(y)\varphi(y)\mathrm{d}y \\ &= \int_{z_1}^{z_2} u(s,y)(\mathcal{L}^*\varphi)(y)\mathrm{d}y - \int_{z_1}^{z_2} H(y)\varphi(y)\mathrm{d}y\end{aligned}$$

where we used integration by parts and $\mathcal{L}^*$ is the adjoint operator of $\mathcal{L}$. Recalling that $u_t \leq 0$ (Proposition 4.4-(i)), we use dominated convergence to obtain

$$\begin{aligned} 0 &\geq \lim_{s\uparrow\bar{c}} \int_{z_1}^{z_2} u(s,y)(\mathcal{L}^*\varphi)(y)\mathrm{d}y - \int_{z_1}^{z_2} H(y)\varphi(y)\mathrm{d}y \\ &= \int_{z_1}^{z_2} \lim_{s\uparrow\bar{c}} u(s,y)(\mathcal{L}^*\varphi)(y)\mathrm{d}y - \int_{z_1}^{z_2} H(y)\varphi(y)\mathrm{d}y \\ &= -\int_{z_1}^{z_2} H(y)\varphi(y)\mathrm{d}y > 0, \end{aligned} \tag{5.12}$$

where the last equality is due to $u(\bar{c},y) = 0$ and the final inequality follows from the facts that $H < 0$ on $(0,\bar{x}_0)$ and φ is arbitrary. Hence a contradiction.

Finally, by the same argument we can prove (5.10). Indeed, if $\mathcal{O} := (0,T)\times(z_1,z_2) \subseteq \mathcal{C}$ for some interval $(z_1,z_2) \subset [0,x_\alpha) \cup (x_\alpha,\bar{x}_0)$. That would imply $c(x) = T$ on (z_1,z_2), contradicting that c cannot be strictly positive and constant. □

Lemma 5.5. *Assume $x_\alpha < \bar{x}_0$. The map $x \mapsto c(x)$ is lower semi-continuous on $\mathbb{R}_+$ and it is continuous at x_α with $c(x_\alpha) = T$. Moreover*

$$c(x) < T \quad \text{for } x \in [0,x_\alpha) \cup (x_\alpha,\bar{x}_0). \tag{5.13}$$

Proof. Recall that $c(x) = T$ for $x \in (\bar{x}_0,\infty)$ by (5.8). Thus $c(\cdot)$ is continuous on $(\bar{x}_0,\infty)$.

Now fix $z \in (0,\bar{x}_0]$ and take a sequence $(z_n)_{n\geq 1} \subseteq (0,\infty)$ with $z_n \to z$ as $n \to \infty$. Then

$$\liminf_{n\to\infty} \big(c(z_n), z_n\big) = \big(\liminf_{n\to\infty} c(z_n), z\big),$$

and since $(c(z_n), z_n)_{n\geq 1} \subseteq \mathcal{S}$ and $\mathcal{S}$ is closed, it must be

$$\big(\liminf_{n\to\infty} c(z_n), z\big) \in \mathcal{S}.$$

The latter implies $\liminf_{n\to\infty} c(z_n) \geq c(z)$, by definition of $c(z)$, and lower semi-continuity follows.

If $z = 0$, then $c(0) = 0$ by Proposition 4.4-(ii). Obviously $\liminf_{n\to\infty} c(z_n) \geq 0$ for any $z_n \to 0$, hence lower semi-continuity holds at $z = 0$ too. To prove continuity at x_α recall that $c(x_\alpha) = T$ by Proposition 5.2, then $\liminf_{n\to\infty} c(z_n) \geq c(x_\alpha) = T$ together with $c(\cdot) \leq T$, imply

$$\liminf_{n\to\infty} c(z_n) = T = \limsup_{n\to\infty} c(z_n)$$

for any $z_n \to x_\alpha$.

It remains to prove (5.13). Since $c(0) = 0$ it suffices to assume there is $x \in (0,\bar{x}_0) \setminus \{x_\alpha\}$ such that $c(x) = T$ and then argue by contradiction. With no loss of generality assume $x \in (x_\alpha,\bar{x}_0)$ since the case of $x \in (0,x_\alpha)$ can be treated analogously. Then we can pick $z_1, z_2 \in (x_\alpha,\bar{x}_0)$ such that $z_1 < x < z_2$ and $\bar{t} := c(z_1) \vee c(z_2) < T$, by (5.10). Hence (ii) in Lemma 5.4 implies $(\bar{t},x) \in \mathcal{S}$, i.e., $c(x) \leq \bar{t}$, which is a contradiction. □

Proof of Proposition 5.3. From (i) in Lemma 5.4 we immediately deduce that $x \mapsto c(x)$ is non-decreasing on $[0,x_\alpha)$. Moreover, (5.13) and the fact that $c(\,\cdot\,)$ cannot be strictly positive and constant (Lemma 5.4) also guarantee that there exists $x_1 \in [0,x_\alpha)$ such that $c(x) = 0$ on $[0,x_1]$ and $c(\,\cdot\,)$ is strictly increasing on $(x_1,x_\alpha]$ (notice that it could be $x_1 = 0$ and $c(\,\cdot\,) > 0$ on $(0,x_\alpha)$). Left-continuity of c on $[0,x_\alpha)$ follows by its monotonicity and lower semi-continuity.

By lower semi-continuity on $\mathbb{R}_+$ and (5.13) there must be a minimum of $c(\,\cdot\,)$ on $[x_\alpha,\bar{x}_0]$, denoted $\hat{c} \in [0,T)$. We have two possible cases: either $\hat{c} = 0$ or $\hat{c} > 0$.

(a) If $\hat{c} = 0$, then the minimum may occur at most on an interval $[x_2, x_3] \subseteq (x_\alpha, \bar{x}_0]$. Indeed, the $\operatorname{argmin}_{[x_\alpha,\bar{x}_0]} c(x)$ is closed by lower semi-continuity of $c(\,\cdot\,)$ and it must be connected by (ii) in Lemma 5.4. However the interval $[x_2, x_3]$ may collapse into a single point $x_2 = x_3$ (in which case $c(\cdot) > 0$ on $[x_\alpha, x_2) \cup (x_2, \bar{x}_0]$);

(b) If $\hat{c} > 0$, then it may only occur at a single point $x_2 (= x_3) \in (x_\alpha, \bar{x}_0]$, again by (ii) in Lemma 5.4 and since $c(\,\cdot\,)$ cannot be strictly positive and constant.

Strict monotonicity on $[x_\alpha, x_2)$ and $(x_3, \bar{x}_0]$ now follows from (ii) in Lemma 5.4 and the fact that $c(\,\cdot\,)$ cannot be strictly positive and constant. Left/right-continuity are then obtained by monotonicity and lower semi-continuity. Finally, the first limit in (5.9) follows by Lemma 5.5, whereas the second one is trivial. □

By arguments as above we obtain analogous results for the case of $x_\alpha \geq \bar{x}_0$. Therefore we omit the proof of the next proposition.

Proposition 5.6. *Assume $x_\alpha \geq \bar{x}_0$. Then, on the interval $[0, \bar{x}_0)$ the map $x \mapsto c(x)$ is non-decreasing, left-continuous, with $c(x) < T$. On the interval $[\bar{x}_0, +\infty)$ it holds $c(x) = T$ and*

$$\lim_{x \uparrow \bar{x}_0} c(x) = c(\bar{x}_0) \leq T. \tag{5.14}$$

Moreover, there exists at most a point $x_1 \leq \bar{x}_0$ such that $c(x) = 0$ for $x \in [0, x_1]$ and $c(\cdot)$ is strictly increasing on $(x_1, \bar{x}_0]$.

5.2. **Higher regularity of the value function and of the optimal boundary.** Thanks to the geometry of the optimal boundary we obtain a lemma that will be used to establish global C^1-regularity of the value function (jointly in (t, x)).

As shown in [14] the key to C^1-regularity of the value function is the probabilistic regularity of the stopping boundary. Since the 2-dimensional process $(t, X_t)_{t \geq 0}$ is not of strong Feller type, we will actually use probabilistic regularity for the *interior* $\mathcal{S}^\circ$ of the stopping region. For completeness we recall that a process $Z \in \mathbb{R}^d$ is said to be of strong Feller type if $z \mapsto \mathsf{E}_z[f(Z_t)]$ is continuous for any $t > 0$ and any *bounded measurable* function $f : \mathbb{R}^d \to \mathbb{R}$.

More precisely, letting

$$\sigma_*(t, x) := \inf\{s \in (0, T - t] : (t + s, X_s^x) \in \mathcal{S}\}, \quad \text{and}$$
$$\sigma_*^\circ(t, x) := \inf\{s \in (0, T - t] : (t + s, X_s^x) \in \mathcal{S}^\circ\}.$$

we say that a boundary point $(t, x) \in \partial\mathcal{C}$ is (probabilistically) regular for $\mathcal{S}$ (or $\mathcal{S}^\circ$) if

$$\mathsf{P}(\sigma_*(t, x) = 0) = 1 \qquad (\text{or } \mathsf{P}(\sigma_*^\circ(t, x) = 0) = 1). \tag{5.15}$$

Clearly, probabilistic regularity for $\mathcal{S}^\circ$ implies the one for $\mathcal{S}$. However, regularity for $\mathcal{S}^\circ$ is meaningless at points $(t_0, z_0) \in \partial\mathcal{C}$ such that $\mathcal{S}$ has empty interior in a neighbourhood of (t_0, z_0). Therefore in what follows we need both.

Lemma 5.7. *The boundary $\partial\mathcal{C}$ is probabilistically regular for $\mathcal{S}$. Moreover, for any $(t_0, z_0) \in \partial\mathcal{C}$ and any sequence $(t_n, x_n) \to (t_0, z_0)$ as $n \to \infty$, it holds*

$$\lim_{n \to \infty} \tau^*(t_n, x_n) = 0, \quad \mathsf{P}\text{-}a.s. \tag{5.16}$$

where τ^* is defined in (3.6).

Proof. By the law of iterated logarithm and the geometry of the stopping region, it is clear that

$$\sigma_*(t, z) = \sigma_*^\circ(t, z) = \tau^*(t, z), \quad \mathsf{P}\text{-a.s.} \tag{5.17}$$

for all $(t,z) \in \partial\mathcal{C}$ except at most along vertical stretches of the boundary corresponding to $x_1 = 0$ and $x_2 = x_3$, as defined in Proposition 5.3. Indeed, at such points a *spike* may occur so that $\mathcal{S}^\circ$ may be (locally) empty. For simplicity let us denote

$$\mathcal{E} := \Big(\big(0, c(x_1+)\big) \times \{x_1\} \Big) \cup \Big(\big(\hat{c}, c(x_2-) \wedge c(x_2+)\big) \times \{x_2\} \Big).$$

By definition $\tau^*(t,z) = 0$, P-a.s., for all $(t,z) \in \partial\mathcal{C}$. Then, by (5.17) we have regularity of $\partial\mathcal{C} \setminus \mathcal{E}$ for $\mathcal{S}^\circ$ in the sense of (5.15). Hence (5.16) holds for any $(t,z) \in \partial\mathcal{C} \setminus \mathcal{E}$ (see, e.g., Corollary 6 in [14]).

Thanks to lower semi-continuity of c, it only remains to consider regularity at $\mathcal{E}$ in the cases: (a) $x_2 = x_3$ but $\hat{c} < c(x_2\pm)$, and (b) $x_1 = 0$ but $c(x_1+) > 0$. We give a full argument for case (a), then case (b) may be handled analogously.

Let us assume $x_2 = x_3$ but $\hat{c} < c(x_2\pm)$. Then $\sigma_*(t,x_2) = \tau^*(t,x_2)$, P-a.s., continues to hold for all $t \in [0,T)$ such that $(t,x_2) \in \partial\mathcal{C}$, by the law of iterated logarithm. Hence the first in (5.15) holds. Since the hitting time $\sigma_*^\circ(t,x_2)$ is no longer zero for $\hat{c} \le t < c(x_2+) \wedge c(x_2-)$, because there is no interior part to the stopping region in a neighbourhood of (t,x_2), the argument provided in [14] to prove the analogue of (5.16) needs a small tweak.

Fix $(t_0,x_2) \in \partial\mathcal{C}$ with $\hat{c} \le t_0 < c(x_2+) \wedge c(x_2-)$ and a sequence $(t_n,x_n)_{n\ge1} \subset \mathcal{C}$ that converges to (t_0,x_2) as $n \to \infty$. Recall that we work with a continuous modification of the stochastic flow and let us pick $\omega \in \Omega$ outside a null set such that $(t,x) \mapsto X_t^x(\omega)$ is continuous. Then for any $\delta > 0$, there exist $0 < s_{1,\omega} < s_{2,\omega} < \delta$ such that $X_{s_{1,\omega}}^{x_2}(\omega) < x_2 < X_{s_{2,\omega}}^{x_2}(\omega)$, by the law of iterated logarithm. By continuity of $x \mapsto X^x$, for some $N_{\delta,\omega} \ge 1$ and all $n \ge N_{\delta,\omega}$, we have $X_{s_{1,\omega}}^{x_n}(\omega) < x_2 < X_{s_{2,\omega}}^{x_n}(\omega)$. With no loss of generality we may assume that $N_{\delta,\omega}$ is sufficiently large that $t_n \ge t_0 - s_{1,\omega}$ for $n \ge N_{\delta,\omega}$. Hence, the points $(t_n + s_{1,\omega}, X_{s_{1,\omega}}^{x_n}(\omega))$ and $(t_n + s_{2,\omega}, X_{s_{2,\omega}}^{x_n}(\omega))$ lie in the two opposite half-planes that are adjacent to the segment $[t_0,T] \times \{x_2\}$. This implies that for each $n \ge N_{\delta,\omega}$ there is $s_{n,\omega} \in (s_{1,\omega}, s_{2,\omega})$ such that $X_{s_{n,\omega}}^{x_n}(\omega) = x_2$ and therefore $(t_n + s_{n,\omega}, X_{s_{n,\omega}}^{x_n}(\omega)) \in \mathcal{S}$. The latter implies $\tau^*(t_n,x_n)(\omega) \le \delta$ for all $n \ge N_{\delta,\omega}$, hence

$$\limsup_{n\to\infty} \tau^*(t_n,x_n)(\omega) \le \delta.$$

Since $\delta > 0$ and ω were arbitrary we obtain (5.16). □

We now provide some useful estimates for $\partial_x v$ in $\mathcal{C}$. Below we use that $(t_0,z_0) \in \partial\mathcal{C}$ with $t_0 < T$ guarantees $z_0 \ne x_\alpha$ by Proposition 5.2, hence $\partial_x h$ is continuous at z_0. In particular, $\partial_x h\,(X_{\tau^*}^x)$ is well-defined on the event $\{\tau^* < T - t\}$, for any $(t,x) \in [0,T) \times \mathbb{R}_+$.

Lemma 5.8. *For all $(t,x) \in \mathcal{C}$ and $0 < s < (\delta^{-1}\ln(2)) \wedge (T-t)$ it holds*

$$\begin{aligned} e^{\delta s}\big(\mathsf{E}\left[\mathbb{1}_{\{\tau^*\le s\}}\partial_x h\,(X_{\tau^*}^x)\right] - \kappa_0 \mathsf{P}\,(\tau^* > s)\big) \\ \le \partial_x v(t,x) \le \big(2 - e^{\delta s}\big)\mathsf{E}\left[\mathbb{1}_{\{\tau^*<s\wedge\tau^\dagger\}}\partial_x h\,(X_{\tau^*}^x)\right], \end{aligned} \tag{5.18}$$

with $\tau^* = \tau^*(t,x)$ and $\tau^\dagger = \tau^\dagger(x)$.

Proof. Recall that for any initial condition $(t,x) \in [0,T] \times \mathbb{R}_+$ the process $V_s^{t,x} = v(t+s, X_s^x)$ is a continuous supermartingale and $s \mapsto V_{s\wedge\tau^*}^{t,x}$ is a continuous martingale for $s \in [0,T-t]$. Fix $(t,x) \in \mathcal{C}$ and $\varepsilon > 0$ such that $(t,x+\varepsilon) \in \mathcal{C}$ and $(t,x-\varepsilon) \in \mathcal{C}$. Notice that

$$\tau^\dagger(x-\varepsilon) \le \tau^\dagger(x) \le \tau^\dagger(x+\varepsilon), \quad \mathsf{P}\text{-a.s.}$$

and, by (ii) in Proposition 4.4, that $\tau^\dagger(x) \geq \tau^*(t,x)$ a.s., because $[0,T]\times\{0\}\subseteq\mathcal{S}$. Set $\tau^* := \tau^*(t,x)$ to simplify notation. Then for all $s < T-t$, using the (super)martingale property, we have

$$\begin{aligned} v(t,x+\varepsilon) &\geq \mathsf{E}\left[v\big(t+(s\wedge\tau^*), X^{x+\varepsilon}_{s\wedge\tau^*}\big)\right] \\ v(t,x) &= \mathsf{E}\left[v\big(t+(s\wedge\tau^*), X^{x}_{s\wedge\tau^*}\big)\right]. \end{aligned}$$

Thus

$$\begin{aligned} & v(t,x+\varepsilon) - v(t,x) \\ & \geq \mathsf{E}\left[v\big(t+(s\wedge\tau^*), X^{x+\varepsilon}_{s\wedge\tau^*}\big) - v\big(t+(s\wedge\tau^*), X^{x}_{s\wedge\tau^*}\big)\right] \\ & = \mathsf{E}\left[\mathbb{1}_{\{\tau^*\leq s\}}\Big(v\big(t+\tau^*, X^{x+\varepsilon}_{\tau^*}\big) - v\big(t+\tau^*, X^{x}_{\tau^*}\big)\Big)\right] \\ & \quad + \mathsf{E}\left[\mathbb{1}_{\{\tau^*> s\}}\Big(v\left(t+s, X^{x+\varepsilon}_{s}\right) - v\left(t+s, X^{x}_{s}\right)\Big)\right] \\ & \geq \mathsf{E}\left[\mathbb{1}_{\{\tau^*\leq s\}}\Big(h\left(X^{x+\varepsilon}_{\tau^*}\right) - h\left(X^{x}_{\tau^*}\right)\Big)\right] - \kappa_0\mathsf{E}\Big[\mathbb{1}_{\{\tau^*> s\}}\big|X^{x+\varepsilon}_{s} - X^{x}_{s}\big|\Big], \end{aligned} \tag{5.19}$$

where $\kappa_0 > 0$ is as in (4.10).

On $\{\tau^*\leq s\}$, the decreasing property of h and $\left|X^{x+\varepsilon}_s - X^x_s\right| \leq \varepsilon e^{\delta s}$ (see (4.1)) give $h\left(X^{x+\varepsilon}_{\tau^*}\right) \geq h\left(X^x_{\tau^*} + \varepsilon e^{\delta s}\right)$. Hence

$$\begin{aligned} v(t,x+\varepsilon) - v(t,x) &\geq \mathsf{E}\left[\mathbb{1}_{\{\tau^*\leq s\}}\left(h\big(X^x_{\tau^*} + \varepsilon e^{\delta s}\big) - h\big(X^x_{\tau^*}\big)\right)\right] - \kappa_0\,\varepsilon\, e^{\delta s}\mathsf{P}\left(\tau^* > s\right) \\ &= \mathsf{E}\left[\mathbb{1}_{\{\tau^*\leq s\}}\int_0^{\varepsilon e^{\delta s}} \partial_x h\left(X^x_{\tau^*} + z\right)\mathrm{d}z\right] - \kappa_0\,\varepsilon\, e^{\delta s}\mathsf{P}\left(\tau^* > s\right), \end{aligned}$$

since h is absolutely continuous on $\mathbb{R}_+$. Then

$$\begin{aligned} \partial_x v(t,x) &= \lim_{\varepsilon\downarrow 0}\frac{1}{\varepsilon}\Big(v(t,x+\varepsilon) - v(t,x)\Big) \\ &\geq \lim_{\varepsilon\downarrow 0}\mathsf{E}\left[\mathbb{1}_{\{\tau^*\leq s\}}\frac{1}{\varepsilon}\int_0^{\varepsilon e^{\delta s}} \partial_x h\left(X^x_{\tau^*} + z\right)\mathrm{d}z\right] - \kappa_0 e^{\delta s}\mathsf{P}\left(\tau^* > s\right) \\ &= \mathsf{E}\left[\mathbb{1}_{\{\tau^*\leq s\}}\lim_{\varepsilon\downarrow 0}\frac{1}{\varepsilon}\int_0^{\varepsilon e^{\delta s}} \partial_x h\left(X^x_{\tau^*} + z\right)\mathrm{d}z\right] - \kappa_0 e^{\delta s}\mathsf{P}\left(\tau^* > s\right), \end{aligned}$$

where the final equality follows by dominated convergence since $\left|\partial_x h\right| \leq 1$.

Now, for each $\omega\in\{\tau^*\leq s\}$ we have $X^x_{\tau^*}(\omega) \neq x_\alpha$ by Proposition 5.2 since $s < T-t$. Hence, there exists $\bar{\varepsilon}_\omega > 0$ such that the mapping $z \mapsto \partial_x h\left(X^x_{\tau^*}(\omega) + z\right)$ is continuous on $\left[0, \bar{\varepsilon}_\omega e^{\delta s}\right]$ and an application of the fundamental theorem of calculus gives

$$\lim_{\varepsilon\downarrow 0}\frac{1}{\varepsilon}\int_0^{\varepsilon e^{\delta s}} \partial_x h\left(X^x_{\tau^*}(\omega) + z\right)\mathrm{d}z = e^{\delta s}\partial_x h\left(X^x_{\tau^*}(\omega)\right). \tag{5.20}$$

Hence

$$\partial_x v(t,x) \geq e^{\delta s}\left(\mathsf{E}\left[\mathbb{1}_{\{\tau^*\leq s\}}\partial_x h\left(X^x_{\tau^*}\right)\right] - \kappa_0\mathsf{P}\left(\tau^* > s\right)\right).$$

Next we want to bound from above the difference $v(t,x) - v(t,x-\varepsilon)$. This requires a slight modification of the previous argument in order to account for the fact that $\tau^\dagger(x-\varepsilon) \leq \tau^\dagger(x)$, a.s. In particular, with no loss of generality we assume that $\varepsilon\in(0,\varepsilon_0]$ for some $\varepsilon_0 > 0$ fixed. Letting $\tau^\dagger_0 := \tau^\dagger(x-\varepsilon_0)$ for simplicity, we have $\tau^\dagger_0 \leq \tau^\dagger(x-\varepsilon)$.

Then, arguing as in (5.19) gives

$$\begin{aligned}
&v(t,x)-v(t,x-\varepsilon)\\
&\le \mathsf{E}\left[v\left(t+(s\wedge\tau^*\wedge\tau_0^\dagger),X^x_{s\wedge\tau^*\wedge\tau_0^\dagger}\right)-v\left(t+(s\wedge\tau^*\wedge\tau_0^\dagger),X^{x-\varepsilon}_{s\wedge\tau^*\wedge\tau_0^\dagger}\right)\right]\\
&\le \mathsf{E}\left[\mathbb{1}_{\{\tau^*\le s\wedge\tau_0^\dagger\}}\left(h\left(X^x_{\tau^*}\right)-h\left(X^{x-\varepsilon}_{\tau^*}\right)\right)\right]\\
&\qquad+\mathsf{E}\left[\mathbb{1}_{\{\tau^*>s\wedge\tau_0^\dagger\}}\left(v\left(t+(s\wedge\tau_0^\dagger),X^x_{s\wedge\tau_0^\dagger}\right)-v\left(t+(s\wedge\tau_0^\dagger),X^{x-\varepsilon}_{s\wedge\tau_0^\dagger}\right)\right)\right].
\end{aligned}$$

Notice that the second term in the last expression is negative thanks to (ii)-Proposition 4.4. Moreover, on the event $\{\tau^*\le s\wedge\tau_0^\dagger\}$ we have

$$h\left(X^x_{\tau^*}\right)-h\left(X^{x-\varepsilon}_{\tau^*}\right)=\int_0^{X^x_{\tau^*}-X^{x-\varepsilon}_{\tau^*}}\partial_x h\left(X^{x-\varepsilon}_{\tau^*}+z\right)\mathrm{d}z\le\int_0^{X^x_{\tau^*}-X^{x-\varepsilon}_{\tau^*}}\partial_x h\left(X^x_{\tau^*}+z\right)\mathrm{d}z$$

where the last step follows from the convexity of $h(\cdot)$. Also, on the event $\{\tau^*\le s\wedge\tau_0^\dagger\}$, using (4.2) we have

$$X^x_{\tau^*}-X^{x-\varepsilon}_{\tau^*}\ge\varepsilon\left(2-e^{\delta\tau^*}\right)\ge\varepsilon\left(2-e^{\delta s}\right)>0,$$

by assuming $s<\delta^{-1}\ln(2)$ with no loss of generality. It follows that, since $\partial_x h\le 0$, we have

$$h\left(X^x_{\tau^*}\right)-h\left(X^{x-\varepsilon}_{\tau^*}\right)\le\int_0^{\varepsilon\left(2-e^{\delta s}\right)}\partial_x h\left(X^x_{\tau^*}+z\right)\mathrm{d}z$$

on $\{\tau^*\le s\wedge\tau_0^\dagger\}$. Thus

$$v(t,x)-v(t,x-\varepsilon)\le\mathsf{E}\left[\mathbb{1}_{\{\tau^*\le s\wedge\tau_0^\dagger\}}\int_0^{\varepsilon\left(2-e^{\delta s}\right)}\partial_x h\left(X^x_{\tau^*}+z\right)\mathrm{d}z\right],$$

and, by arguments as in (5.20), we obtain

$$\partial_x v(t,x)\le\left(2-e^{\delta s}\right)\mathsf{E}\left[\mathbb{1}_{\{\tau^*\le s\wedge\tau_0^\dagger\}}\partial_x h\left(X^x_{\tau^*}\right)\right],\qquad\text{for } s<(\delta^{-1}\ln(2))\wedge(T-t).$$

To conclude we let $\varepsilon_0\downarrow 0$ so that $\tau_0^\dagger=\tau^\dagger(x-\varepsilon_0)\uparrow\tau^\dagger(x)$ and the upper bound in (5.18) holds by monotone convergence. □

Proposition 5.9. *Fix any $(t_0,z_0)\in\partial\mathcal{C}$ with $t_0<T$ and $z_0>0$. Then, for any sequence $(t_n,x_n)_{n\ge1}\subset\mathcal{C}$ such that $(t_n,x_n)\to(t_0,z_0)$ as $n\uparrow\infty$, we have*

$$\lim_{n\to\infty}\partial_x v(t_n,x_n)=\partial_x h(z_0)\tag{5.21}$$

and

$$\lim_{n\to\infty}\partial_t v(t_n,x_n)=0.\tag{5.22}$$

Proof. First we prove (5.21). Notice, that (5.18) holds for any point $(t_n,x_n)\in\mathcal{C}$ from a sequence that converges to (t_0,z_0) as $n\uparrow\infty$, where $(t_0,z_0)\in\partial\mathcal{C}$ with $t_0<T$ and $z_0>0$. Since $\tau_n^*:=\tau^*(t_n,x_n)\to 0$ as $n\uparrow\infty$, by Lemma 5.7, and $\partial_x h\left(X^{x_n}_{\tau_n^*}\right)\to\partial_x h(z_0)$ (recall that $z_0\ne x_\alpha$), then for $0<s<(\delta^{-1}\ln(2))\wedge(T-t_0)$ dominated convergence and (5.18) give

$$e^{\delta s}\partial_x h(z_0)\le\liminf_{n\to\infty}\partial_x v(t_n,x_n)\le\limsup_{n\to\infty}\partial_x v(t_n,x_n)\le(2-e^{\delta s})\partial_x h(z_0).$$

Letting $s\to 0$ we get (5.21).

To prove (5.22), fix $(t,x)\in\mathcal{C}$ with $t<T$ and $\varepsilon>0$ such that $(t+\varepsilon,x)\in\mathcal{C}$. Let

$$\tau_N := \inf\{u\geq 0 : X^x_u \geq N\},$$

and pick $s<T-(t+\varepsilon)$. Then, proceeding as in the proof of Lemma 5.8, with $\tau^*=\tau^*(t,x)$ we have

$$\begin{aligned} v(t+\varepsilon,x) &\geq \mathsf{E}\left[v\big(t+\varepsilon+(s\wedge\tau^*\wedge\tau_N), X^x_{s\wedge\tau^*\wedge\tau_N}\big)\right], \\ v(t,x) &= \mathsf{E}\left[v\big(t+(s\wedge\tau^*\wedge\tau_N), X^x_{s\wedge\tau^*\wedge\tau_N}\big)\right]. \end{aligned}$$

Combining the above with (4.11) gives

$$\begin{aligned} &v(t+\varepsilon,x)-v(t,x) \\ &\geq \mathsf{E}\left[\mathbb{1}_{\{\tau^*\leq s\wedge\tau_N\}}\Big(h\left(X^x_{\tau^*}\right)-h\left(X^x_{\tau^*}\right)\Big)\right] \\ &\qquad + \mathsf{E}\left[\mathbb{1}_{\{\tau^*>s\wedge\tau_N\}}\Big(v\big(t+\varepsilon+(s\wedge\tau_N), X^x_{s\wedge\tau_N}\big)-v\big(t+(s\wedge\tau_N), X^x_{s\wedge\tau_N}\big)\Big)\right] \\ &\geq -\kappa_1(t+\varepsilon+s,N)\,\varepsilon\,\mathsf{P}\left(\tau^*>s\wedge\tau_N\right). \end{aligned}$$

With no loss of generality we may assume that $\varepsilon\in(0,\varepsilon_0]$ for some $\varepsilon_0>0$ such that $s<T-(t+\varepsilon_0)$ and $\kappa_1(t+\varepsilon+s,N)\leq\hat{\kappa}_1(\varepsilon_0,N)$ for some constant $\hat{\kappa}_1(\varepsilon_0,N)>0$. Hence,

$$0\geq\partial_t v(t,x)\geq -\hat{\kappa}_1(\varepsilon_0,N)\,\mathsf{P}\left(\tau^*>s\wedge\tau_N\right).$$

The result holds for any $(t_n,x_n)\in\mathcal{C}$ from a sequence converging to (t_0,x_0). Moreover, with no loss of generality we can assume that $x_n\leq x_0+1<N$ for all $n\geq 1$ so that $\tau_N(x_n)\geq\tau_N(x_0+1)>0$, P-a.s., and

$$\mathsf{P}\Big(\tau^*(t_n,x_n)>s\wedge\tau_N(x_n)\Big)\leq\mathsf{P}\Big(\tau^*(t_n,x_n)>s\wedge\tau_N(x_0+1)\Big).$$

Then, thanks to Lemma 5.7 we get

$$0\geq\limsup_{n\to\infty}\partial_t v(t_n,x_n)\geq\liminf_{n\to\infty}\partial_t v(t_n,x_n)\geq 0.$$

Hence (5.22) holds. □

Proposition 5.9 and Proposition 4.6 imply continuous differentiability of v.

Corollary 5.10. *The value function v is continuously differentiable on the set $[0,T)\times(0,+\infty)$. Moreover, $v\in C^{1,2}\big(\overline{\mathcal{C}}\cap\big([0,T)\times(0,+\infty)\big)\big)$ with*

$$\lim_{\mathcal{C}\ni(t,x)\to(t_0,z_0)\in\partial\mathcal{C}}\partial_{xx}v(t,x)=-\frac{2}{\sigma^2}\pi(z_0)\partial_x h(z_0) \tag{5.23}$$

for $t_0<T$ and $z_0>0$.

Proof. We only need to prove (5.23). In order to do that it is sufficient to take limits in (4.15) and use Proposition 5.9. □

The next theorem shows that the optimal boundary is continuous as a function of x. Notice that this type of continuity is not a standard result in optimal stopping problems for time-space processes (t,X). Indeed, in the probabilistic literature, one normally proves continuity of the boundary as a function of *time*. Our proof relies on the use of a suitably constructed reflecting diffusion.

Theorem 5.11. *The mapping $x\mapsto c(x)$ is continuous on $(0,\infty)$. If $c(0+)=0$ then continuity holds on $\mathbb{R}_+$ (recall $\mathbb{R}_+=[0,\infty)$).*

Proof. We give a full proof in the case $x_\alpha < \bar{x}_0$ and consider the interval $[x_3, \infty)$, with x_3 as in Proposition 5.3, where the boundary is increasing. It will be clear that the intervals $[x_1, x_\alpha]$ and $[x_\alpha, x_2]$ and the case $x_\alpha \geq \bar{x}_0$ can be treated analogously.

Arguing by contradiction let us assume that there exists $z_0 \in [x_3, \infty)$ such that $c(z_0) < c(z_0+)$ and let $\mathcal{I}_0 := (c(z_0), c(z_0+))$. Then $\mathcal{I}_0 \times \{z_0\} \subset \partial\mathcal{C}$ and there exists $z_1 > z_0$ such that $\partial_t u(t, z_1) < -\varepsilon_1$ for some $\varepsilon_1 > 0$ and for all $t \in \mathcal{I}_1 := (t_0, t_1) \subset \mathcal{I}_0$ for some $t_1 > t_0$ (see Lemma 4.7).

Since $u \in C^1([0,T) \times \mathbb{R}_+)$ by Corollary 5.10 and $\mathcal{I}_1 \times \{z_0\} \subset \partial\mathcal{C}$, we have $\partial_t u(t, z_0) = \partial_x u(t, z_0) = 0$ for $t \in \mathcal{I}_1$. Then for any $\varepsilon > 0$ there exists $\delta_\varepsilon > 0$ such that $z_0 + \delta_\varepsilon < z_1$ and

$$0 \geq \partial_t u \geq -\varepsilon \quad \text{and} \quad |\partial_x u| \leq \varepsilon \quad \text{on } \overline{\mathcal{I}_1} \times [z_0, z_0 + \delta_\varepsilon], \tag{5.24}$$

by uniform continuity on any compact.

Now we consider a process that equals $(X_t)_{t\geq 0}$ away from $z_0 + \delta_\varepsilon$ and is reflected (upwards) at $z_0 + \delta_\varepsilon$. It is well-known (see, e.g., [27] or [4, Sec. 12, Chapter I]) that there exists a unique strong solution of the stochastic differential equation

$$\mathrm{d}\widetilde{X}^\varepsilon_t = \pi(\widetilde{X}^\varepsilon_t)\mathrm{d}t + \sigma \mathrm{d}W_t + \mathrm{d}A^{\delta_\varepsilon}_t, \qquad \widetilde{X}^\varepsilon_0 = z_0 + \delta_\varepsilon,$$

where A^{δ_ε} is a continuous, non-decreasing process that guarantees, P-a.s.,

$$\widetilde{X}^\varepsilon_t \geq z_0 + \delta_\varepsilon, \quad \text{for all } t \geq 0 \text{ and} \qquad \int_0^T \mathbb{1}_{\{\widetilde{X}^\varepsilon_t > z_0 + \delta\}} \mathrm{d}A^{\delta_\varepsilon}_t = 0. \tag{5.25}$$

As in Lemma 4.7 we appeal to classical results on interior regularity for solutions of PDEs that guarantee $\partial_t u \in C^{1,2}\big(\mathcal{I}_1 \times (z_0, z_1)\big)$ and $(\partial_t + \mathcal{L})\partial_t u = 0$ on $\mathcal{I}_1 \times (z_0, z_1)$. Then, setting $\tau^\varepsilon_1 := \inf\{s \geq 0 : \widetilde{X}^\varepsilon_s = z_1\}$ and $\hat{u} := \partial_t u$, an application of Itô's formula for semi-martingales gives, for any $t \in \mathcal{I}_1$

$$\begin{aligned} &\mathsf{E}\Big[\hat{u}(t + \tau^\varepsilon_1 \wedge (t_1 - t), \widetilde{X}^\varepsilon_{\tau^\varepsilon_1 \wedge (t_1 - t)})\Big] \\ =&\hat{u}(t, z_0 + \delta_\varepsilon) + \mathsf{E}\left[\int_0^{\tau^\varepsilon_1 \wedge (t_1 - t)} \partial_x \hat{u}(t + s, \widetilde{X}^\varepsilon_s)\, \mathrm{d}A^{\delta_\varepsilon}_s\right] \\ \geq& -\varepsilon + \mathsf{E}\left[\int_0^{\tau^\varepsilon_1 \wedge (t_1 - t)} \partial_{tx} u(t + s, z_0 + \delta_\varepsilon)\, \mathrm{d}A^{\delta_\varepsilon}_s\right] \end{aligned} \tag{5.26}$$

where the inequality follows from (5.24) and the second condition in (5.25) implies

$$\mathrm{d}A^{\delta_\varepsilon}_s = \mathbb{1}_{\{\widetilde{X}^\varepsilon_s = z_0 + \delta_\varepsilon\}} \mathrm{d}A^{\delta_\varepsilon}_s.$$

For the expression on the left-hand side of (5.26) we have

$$\begin{aligned} &\mathsf{E}\left[\hat{u}(t + \tau^\varepsilon_1 \wedge (t_1 - t), \widetilde{X}^\varepsilon_{\tau^\varepsilon_1 \wedge (t_1 - t)})\right] \\ \leq&\mathsf{E}\left[\mathbb{1}_{\{\tau^\varepsilon_1 < t_1 - t\}} \hat{u}(t + \tau^\varepsilon_1, z_1)\right] \leq -\varepsilon_1\, \mathsf{P}(\tau^\varepsilon_1 < t_1 - t). \end{aligned}$$

Hence, from (5.26) we obtain

$$-\varepsilon_1 \mathsf{P}(\tau^\varepsilon_1 < t_1 - t) \geq -\varepsilon + \mathsf{E}\left[\int_0^{\tau^\varepsilon_1 \wedge (t_1 - t)} \partial_{tx} u(t + s, z_0 + \delta_\varepsilon)\, \mathrm{d}A^{\delta_\varepsilon}_s\right]. \tag{5.27}$$

The next step is to let $\varepsilon \to 0$. However, the regularity of $\partial_{tx} u$ as $\delta_\varepsilon \downarrow 0$ might be problematic. We therefore use a trick with test functions to overcome this difficulty.

Pick $\varphi \in C_c^\infty(\mathcal{I}_1)$, $\varphi \geq 0$ such that $\int_{\mathcal{I}_1} \varphi(t)\, dt = 1$. Then, multiplying both sides of (5.27) by φ, integrating over $\mathcal{I}_1$ and using Fubini's theorem we obtain

$$\begin{aligned}
&-\varepsilon_1 \int_{\mathcal{I}_1} \mathsf{P}(\tau_1^\varepsilon < t_1 - t)\varphi(t)\mathrm{d}t \\
&\quad \geq -\varepsilon + \mathsf{E}\left[\int_0^{\tau_1^\varepsilon} \left(\int_{\mathcal{I}_1} \mathbb{1}_{\{t<t_1-s\}} \partial_{tx} u(t+s, z_0+\delta_\varepsilon)\varphi(t)\mathrm{d}t\right) \mathrm{d}A_s^{\delta_\varepsilon}\right] \\
&\quad = -\varepsilon + \mathsf{E}\Bigg[\int_0^{\tau_1^\varepsilon} \Bigg(\partial_x u(t_1, z_0+\delta_\varepsilon)\varphi(t_1-s) \\
&\qquad\qquad - \int_{\mathcal{I}_1} \mathbb{1}_{\{t<t_1-s\}} \partial_x u(t+s, z_0+\delta_\varepsilon)\varphi'(t)\mathrm{d}t\Bigg)\mathrm{d}A_s^{\delta_\varepsilon}\Bigg] \\
&\quad \geq -\varepsilon - \varepsilon\, \mathsf{E}\Bigg[\int_0^{\tau_1^\varepsilon} \varphi(t_1-s)\mathrm{d}A_s^{\delta_\varepsilon} - A^{\delta_\varepsilon}_{\tau_1^\varepsilon\wedge t_1} \int_{\mathcal{I}_1} |\varphi'(t)|\mathrm{d}t\Bigg] \\
&\quad \geq -\varepsilon\left(1 + \left(\|\varphi\|_\infty + T\|\varphi'\|_\infty\right)\mathsf{E}\big[A^{\delta_\varepsilon}_{\tau_1^\varepsilon\wedge t_1}\big]\right),
\end{aligned}$$

where for the penultimate inequality we have used the bounds on $\partial_x u$ given in (5.24), and the final inequality uses that $\varphi(t_1 - s) = 0$ for $s \geq t_1$. Here $\|\cdot\|_\infty$ is the supremum norm on $[0, T]$.

For the increasing process A^{δ_ε} we have an upper bound which is independent of ε. This can be deduced from the integral form of the SDE. That is, taking expectation of

$$\widetilde{X}^\varepsilon_{\tau_1^\varepsilon\wedge t_1} = z_0 + \delta_\varepsilon + \int_0^{\tau_1^\varepsilon\wedge t_1} \pi(\widetilde{X}^\varepsilon_s)\mathrm{d}s + \sigma W_{\tau_1^\varepsilon\wedge t_1} + A^{\delta_\varepsilon}_{\tau_1^\varepsilon\wedge t_1}$$

gives

$$\mathsf{E}\left[A^{\delta_\varepsilon}_{\tau_1^\varepsilon\wedge t_1}\right] = \mathsf{E}\left[\widetilde{X}^\varepsilon_{\tau_1^\varepsilon\wedge t_1} - z_0 - \delta_\varepsilon - \int_0^{\tau_1^\varepsilon\wedge t_1} \pi(\widetilde{X}^\varepsilon_s)\mathrm{d}s\right].$$

Using that $\widetilde{X}^\varepsilon_s \in [z_0 + \delta_\varepsilon, z_1]$ and $\pi(\widetilde{X}^\varepsilon_s) \geq \pi(z_1)$ for $s \leq \tau_1^\varepsilon$, we obtain

$$\mathsf{E}\left[A^{\delta_\varepsilon}_{\tau_1^\varepsilon\wedge t_1}\right] \leq c_0 := z_1 - z_0 - t_1\, \pi(z_1).$$

Hence

$$-\varepsilon_1 \int_{\mathcal{I}_1} \mathsf{P}(\tau_1^\varepsilon < t_1 - t)\varphi(t)\mathrm{d}t \geq -\varepsilon\left(1 + \left(\|\varphi\|_\infty + \|\varphi'\|_\infty\right) c_0\right)$$

and taking limits as $\varepsilon \to 0$ gives

$$\limsup_{\varepsilon\to 0} \int_{\mathcal{I}_1} \mathsf{P}(\tau_1^\varepsilon < t_1 - t)\varphi(t)\mathrm{d}t \leq 0. \tag{5.28}$$

If we can show that the left hand side above is positive we have reached a contradiction, and there cannot be a discontinuity of c.

For our final task we introduce the change of measure

$$\left.\frac{\mathrm{d}\widehat{\mathsf{P}}^\varepsilon}{\mathrm{d}\mathsf{P}}\right|_{\mathcal{F}_T} := \exp\left(-\int_0^T \sigma^{-1}\pi(\widetilde{X}^\varepsilon_s)\mathrm{d}W_s - \frac{1}{2}\int_0^T \sigma^{-2}\pi^2(\widetilde{X}^\varepsilon_s)\mathrm{d}s\right), \tag{5.29}$$

so that under $\widehat{\mathsf{P}}^\varepsilon$, we have

$$\widetilde{X}^\varepsilon_t = z_0 + \delta_\varepsilon + \sigma\widehat{W}^\varepsilon_t + A^{\delta_\varepsilon}_t,$$

where $\widehat{W}^\varepsilon := (\widehat{W}^\varepsilon_t)_{t\in[0,T]}$ is a Brownian motion defined as

$$\widehat{W}^\varepsilon_t = W_t + \int_0^t \sigma^{-1}\pi(\widetilde{X}^\varepsilon_s)\mathrm{d}s.$$

For future reference we also introduce the Dóleans-Dade exponential

$$Z^\varepsilon_T := \exp\left(\int_0^T \sigma^{-1}\pi(\widetilde{X}^\varepsilon_s)\mathrm{d}\widehat{W}^\varepsilon_s - \frac{1}{2}\int_0^T \sigma^{-2}\pi^2(\widetilde{X}^\varepsilon_s)\mathrm{d}s\right). \tag{5.30}$$

Since the measures are equivalent on $\mathcal{F}_T$, under $\widehat{\mathsf{P}}^\varepsilon$ the process $\widetilde{X}^\varepsilon$ is a Brownian motion reflected at $z_0+\delta_\varepsilon$. Hence, we have an explicit formula for the increasing process A^{δ_ε} (see, [24, Lemma 6.14, Chapter 3]), that is

$$A^{\delta_\varepsilon}_t = \sup_{0\le s\le t}\big(-\sigma\widehat{W}^\varepsilon_s\big). \tag{5.31}$$

It remains to remove the dependence of the measure on ε. For that we can take a filtered probability space $(\widehat{\Omega},\widehat{\mathcal{F}},(\widehat{\mathcal{F}}_t)_{t\in[0,T]},\widehat{\mathsf{P}})$ equipped with a standard Brownian motion $B := (B_t)_{t\in[0,T]}$. On such space we construct a Brownian motion starting from $z_0+\delta_\varepsilon$ and reflected at its starting point, that we denote Y. That is

$$Y^\varepsilon_t = z_0+\delta_\varepsilon+\sigma B_t + L_t, \qquad t\in[0,T],$$

where L takes the same expression of (5.31) but with B instead of $\widehat{W}^\varepsilon$. For future reference we also denote

$$Y^0_t = z_0+\sigma B_t + L_t, \qquad t\in[0,T].$$

By construction

$$\mathsf{Law}(\widetilde{X}^\varepsilon\,|\,\widehat{\mathsf{P}}^\varepsilon) = \mathsf{Law}(Y^\varepsilon\,|\,\widehat{\mathsf{P}}).$$

Then, setting $\rho^\varepsilon_1 := \inf\{t\ge 0 : Y^\varepsilon_t = z_1\}$, denoting $\widehat{\mathsf{E}}^\varepsilon$ the expectation under $\widehat{\mathsf{P}}^\varepsilon$ and letting ξ^ε_T be defined as the Dóleans-Dade exponential in (5.30) but with (Y^ε, B) instead of $(\widetilde{X}^\varepsilon,\widehat{W}^\varepsilon)$, we obtain

$$\mathsf{P}(\tau^\varepsilon_1 < t_1 - t) = \widehat{\mathsf{E}}^\varepsilon\left[Z^\varepsilon_T \mathbb{1}_{\{\tau^\varepsilon_1<t_1-t\}}\right] = \widehat{\mathsf{E}}\left[\xi^\varepsilon_T \mathbb{1}_{\{\rho^\varepsilon_1<t_1-t\}}\right]. \tag{5.32}$$

Using the explicit form of Y^ε, under $\widehat{\mathsf{P}}$ we have

$$\lim_{\varepsilon\to 0}\rho^\varepsilon_1 = \rho^0_1 := \inf\{s\ge 0 : z_0+\sigma B_t + L_t = z_1\},$$

where the convergence is monotonic from above and therefore also

$$\lim_{\varepsilon\to 0}\mathbb{1}_{\{\rho^\varepsilon_1<t_1-t\}} = \mathbb{1}_{\{\rho^0_1<t_1-t\}}.$$

Hence, Fatou's lemma and (5.32) give

$$\liminf_{\varepsilon\downarrow 0}\mathsf{P}(\tau^\varepsilon_1 < t_1-t) \ge \widehat{\mathsf{E}}\left[\liminf_{\varepsilon\to 0}\xi^\varepsilon_T\mathbb{1}_{\{\rho^\varepsilon_1<t_1-t\}}\right] = \widehat{\mathsf{E}}\left[\xi^0_T\mathbb{1}_{\{\rho^0_1<t_1-t\}}\right] > 0,$$

where ξ^0_T is the Dóleans-Dade exponential associated to (Y^0, B), and the final inequality follows from well-known distributional properties of reflected Brownian motion (see, e.g., [24, Sec. 2.8.B]).

Finally, using Fatou's lemma in (5.28), and the discussion above, we conclude

$$0 \ge \liminf_{\varepsilon\downarrow 0}\int_{\mathcal{I}_1}\mathsf{P}(\tau^\varepsilon_1<t_1-t)\varphi(t)\mathrm{d}t \ge \int_{\mathcal{I}_1}\widehat{\mathsf{E}}\left[\xi^0_T\mathbb{1}_{\{\rho^0_1<t_1-t\}}\right]\varphi(t)\mathrm{d}t > 0,$$

where the final inequality uses that $\varphi\ge 0$ and arbitrary. Hence a contradiction and continuity of $x\mapsto c(x)$ is proved. □

If $c(0+) > 0$ the smooth-fit may break down on $\{0\} \times \big[0, c(0+)\big)$, i.e., $\partial_x u(t, 0+) \neq 0$ for some $t \in \big[0, c(0+)\big)$. That is why continuity of $c(\,\cdot\,)$ only holds on $(0, \infty)$ in that case. Combining the continuity result with (5.9) and (5.14) also guarantees:

Corollary 5.12. *It holds*

$$\lim_{x \to \bar{x}_0} c(x) = T.$$

We conclude the section by giving the proofs of the main results stated in Section 3.

5.3. Proofs of Theorems 3.1 and 3.2.

Proof of Theorem 3.1. The first claim is consequence of Corollary 4.5, and $v \geq h$ follows by taking $\tau = 0$ in (3.1). Monotonicity of the mappings $t \mapsto v(t, x)$ and $x \mapsto v(t, x)$ was proven in Proposition 4.4. Continuous differentiability of v and continuity of $\partial_{xx} v$ on $\overline{\mathcal{C}} \cap \big([0, T) \times (0, \infty)\big)$ were obtained in Corollary 5.10. Proposition 4.6 guarantees that v solves (3.7) in $\mathcal{C}$. Moreover, $v = h$ in $\mathcal{S}$ and $\mathcal{L}h(x) \leq 0$ in $\mathcal{S}$ by (5.3). Hence (3.7) holds.

As for uniqueness, if we can find another function w that solves (3.7) with $\mathcal{C} = \{w > h\}$ and with the same regularity as v, then by a standard verification argument based on a well-known generalisation of Itô's formula we obtain that w coincides with the value function of the optimal stopping problem (3.1). Further details in this direction are omitted as they are standard and can be found in [6, Thm. 4.2, Ch. IV]. □

Proof of Theorem 3.2. We only provide the full argument for (b), which builds on the results of Proposition 5.3. The proof of (a) is easier and follows from analogous arguments and Proposition 5.6.

If $x_\alpha < \bar{x}_0$ our boundary c is strictly monotonic and continuous on the intervals $[x_1, x_\alpha)$, (x_α, x_2) and $(x_3, \bar{x}_0)$ (with x_1, x_2 and x_3 as in Proposition 5.3) and $c(x) = T$ for $x \geq \bar{x}_0$. Hence, the map $x \mapsto c(x)$ can be inverted separately on the intervals $[x_1, x_\alpha)$, (x_α, x_2) and $(x_3, \bar{x}_0]$, to obtain three continuous and strictly monotonic functions of time that describe the boundary $\partial\mathcal{C}$.

Recalling $\hat{c}$ from Proposition 5.3, we can define locally the inverse functions

$$b_1(t) := \inf\{x \in [0, x_\alpha) : c(x) > t\}, \qquad t \in [0, T), \tag{5.33}$$

$$b_2(t) := \sup\{x \in (x_\alpha, x_2) : c(x) > t\}, \qquad t \in [\hat{c}, T), \tag{5.34}$$

$$b_3(t) := \inf\{x \in (x_3, \bar{x}_0) : c(x) > t\}, \qquad t \in [\hat{c}, T). \tag{5.35}$$

Clearly $0 \leq b_1(t) \leq b_2(t) \leq b_3(t)$ for $t \in [\hat{c}, T)$. Setting $t_0 := c(0+)$ we have $b_1(t) = 0$ on $[0, t_0)$ if $t_0 > 0$, otherwise $b(0+) = x_1$. Recalling that $x \mapsto c(x)$ is strictly increasing and continuous on $[x_1, x_\alpha]$ and $[x_3, \bar{x}_0]$, we obtain that $t \mapsto b_1(t)$ and $t \mapsto b_3(t)$ are strictly increasing and continuous on $[t_0, T)$ and $[\hat{c}, T)$ respectively. Analogously, since $x \mapsto c(x)$ is strictly decreasing and continuous on $[x_\alpha, x_2]$ we have $t \mapsto b_2(t)$ strictly decreasing and continuous on $[\hat{c}, T)$. Moreover, $b_1(T-) = b_2(T-) = x_\alpha$ by the first limit in (5.9) and $b_3(T-) = \bar{x}_0$ by Corollary 5.12. By construction, if $\hat{c} > 0$ we have $b_2(\hat{c}) = b_3(\hat{c})$ because $\hat{c}$ is the unique point in $\operatorname{argmin}_{[x_\alpha, \bar{x}_0]} c(x)$. If instead $\hat{c} = 0$, we have $b_2(0) = x_2$ and $b_3(0) = x_3$ by Proposition 5.3. This concludes the proof of i)–iv).

The claims in v) are now straightforward. □

6. Some comments about management fees

In a model with proportional management fees as in (3.2) the only change in our analysis is due to the fact that the function H appearing in u (see (4.7) and (4.9)) is

replaced by

$$H^{p,q}(x) := \begin{cases} e^{-x}\left(\frac{\sigma^2}{2} - q - \pi(x)\right) - p, & x \le x_\alpha, \\ (1-\gamma)\, e^{-x}\left(\frac{\sigma^2}{2} - \frac{q}{1-\gamma} - \pi(x)\right) - p, & x > x_\alpha, \end{cases}$$

where $p, q \ge 0$ are as in (3.2).

The key features in our analysis from the previous sections are the presence of the local time at x_α, in the problem formulation (4.9), and the sign of the function H via (5.3) (whose interplay leads to the two separate cases $x_\alpha < \bar{x}_0$ and $x_\alpha \ge \bar{x}_0$). Here, the local time produces the same effects, so that Proposition 5.2 continues to hold and $[0,T) \times \{x_\alpha\} \subset \mathcal{C}$. Instead, we now need to look at the sign of the function $H^{p,q}$ rather than that of H.

It is immediate to check that, for $x \le x_\alpha$,

$$H^{p,q}(x) > 0 \implies \frac{\sigma^2}{2} - q - \pi(x) > 0 \iff x > \bar{x}_q := \beta + \frac{r+q}{\delta}$$

and, if $x_\alpha < \bar{x}_q$, then $H^{p,q}(x) < 0$. Similarly, for $x > x_\alpha$,

$$H^{p,q}(x) > 0 \implies \frac{\sigma^2}{2} - \frac{q}{1-\gamma} - \pi(x) > 0 \iff x > \bar{x}_{q,\gamma} := \beta + \frac{r + q(1-\gamma)^{-1}}{\delta},$$

where $\bar{x}_{q,\gamma} > \bar{x}_q > x_G$ since $\gamma \in (0,1)$ (recall $x_G = \beta + r^G/\delta$). Hence, if $x_\alpha < \bar{x}_q$, it is $H^{p,q}(x) < 0$ for $x_\alpha < x < \bar{x}_{q,\gamma}$.

So, also in the presence of management fees we must consider various cases depending on the position of x_α relative to $\bar{x}_q$ and $\bar{x}_{q,\gamma}$. In keeping with the rest of the paper and in the interest of length, here we briefly illustrate only the case $x_\alpha < \bar{x}_q$ and draw a parallel with the case $x_\alpha < \bar{x}_0$ from the previous sections. The remaining cases can be studied analogously with the methods developed above.

From now on, let us assume $x_\alpha < \bar{x}_q$. Then

$$H^{p,q}(x) > 0 \iff x - \bar{x}_{q,\gamma} > \frac{p}{\delta}(1-\gamma)^{-1}e^x. \tag{6.1}$$

In the special case $p = 0$, we have the exact analogue of (b) in Theorem 3.2 but with $\bar{x}_0$ therein, replaced by $\bar{x}_{q,\gamma}$. If instead $p > 0$ only two sub-cases may arise and we must consider them separately:

Case (i): $x - \bar{x}_{q,\gamma} \le \frac{p}{\delta}(1-\gamma)^{-1}e^x$ for all $x \in \mathbb{R}_+$ (that is $\ln[\frac{\delta}{p}(1-\gamma)] \le 1 + \bar{x}_{q,\gamma}$). Then

$$\mathcal{S} \subseteq \Big\{[0,T) \times \Big([0,x_\alpha) \cup (x_\alpha, \infty)\Big)\Big\} \cup \Big(\{T\} \times \mathbb{R}_+\Big);$$

Case (ii): There exist $\hat{x}_2 > \hat{x}_1 > \bar{x}_{q,\gamma}$ such that $x - \bar{x}_{q,\gamma} > \frac{p}{\delta}(1-\gamma)^{-1}e^x$ for all $x \in (\hat{x}_1, \hat{x}_2)$ and $x - \bar{x}_{q,\gamma} \le \frac{p}{\delta}(1-\gamma)^{-1}e^x$ otherwise (that is $\ln[\frac{\delta}{p}(1-\gamma)] > 1 + \bar{x}_{q,\gamma}$). Then

$$\mathcal{S} \subseteq \Big\{[0,T) \times \Big([0,x_\alpha) \cup (x_\alpha, \hat{x}_1] \cup [\hat{x}_2, \infty)\Big)\Big\} \cup \Big(\{T\} \times \mathbb{R}_+\Big).$$

In case (ii) the situation is similar to (5.5), with $\bar{x}_0$ therein replaced by $\hat{x}_1$ and noting the additional strip $[0,T] \times [\hat{x}_2, \infty)$ intersecting the stopping set. In particular, repeating the same arguments as in Proposition 5.3 one can prove that the map $x \mapsto c(x)$ (defined as in (5.7)) satisfies all claims in the proposition with $\bar{x}_0$ replaced by $\hat{x}_1$ throughout. In addition to that, and by the same methods, one can also prove that $c(x) < T$ for all $x > \hat{x}_2$ and there exists $\hat{x}_3 > \hat{x}_2$ such that $x \mapsto c(x)$ is strictly decreasing on $(\hat{x}_2, \hat{x}_3)$

with $c(\hat{x}_2+) = T$ and $c(x) = 0$ for $x \in [\hat{x}_3, \infty)$. Thanks to piece-wise monotonicity of the boundary, also in this setting we can prove its continuity as in Theorem 5.11.

While the strict monotonicity follows by the exact same arguments as those used in the proof of Proposition 5.3, for completeness we prove the existence of $\hat{x}_3$, which did not appear in the previous analysis.

Proposition 6.1. *In the setting of* Case (ii)*, there exists* $\hat{x}_3 > \hat{x}_2$ *such that* $c(x) = 0$ *for* $x \in [\hat{x}_3, \infty)$.

Proof. Let us argue by contradiction and assume $c(x) > 0$ for all $x \geq \hat{x}_2$. In particular, let us first assume the stronger requirement that there exists $\theta > 0$ such that $c(x) \geq \theta$ for all $x \geq \hat{x}_2$. Consider the value function $v(0, x)$ for $x > m > \hat{x}_2$ and a fixed m. Setting $\tau_m = \inf\{t \geq 0 : X_t^x \leq m\}$ and letting τ^* be optimal for $v(0, x)$, we have $\tau^* \geq \theta \wedge \tau_m$, P-a.s. Now, using this observation we have

$$\begin{aligned} v(0,x) =& \mathsf{E}_x\Big[h(X_{\tau^*\wedge\tau^\dagger}) - \int_0^{\tau^*\wedge\tau^\dagger}\Big(p + qe^{-X_t}\Big)\mathrm{d}t\Big] \\ =& \mathsf{E}_x\Big[\mathbb{1}_{\{\tau^*<\tau_m\}}\Big(h(X_{\tau^*\wedge\tau^\dagger}) - \int_0^{\tau^*\wedge\tau^\dagger}\Big(p + qe^{-X_t}\Big)\mathrm{d}t\Big)\Big] \\ &+ \mathsf{E}_x\Big[\mathbb{1}_{\{\tau^*\geq\tau_m\}}\Big(h(X_{\tau^*\wedge\tau^\dagger}) - \int_0^{\tau^*\wedge\tau^\dagger}\Big(p + qe^{-X_t}\Big)\mathrm{d}t\Big)\Big]. \end{aligned}$$

On the event $\{\tau^* < \tau_m\}$ we have $\tau^* < \tau^\dagger$ and $\tau^* \geq \theta$. Moreover, on that event $h(X^x_{\tau^*\wedge\tau^\dagger}) \leq h(m)$ since h is decreasing (see (2.16)) and $X^x_{\tau^*\wedge\tau^\dagger} \geq m$. Recalling that h is positive and bounded by 1, we obtain the upper bound

$$\begin{aligned} v(0,x) \leq& \mathsf{E}_x\Big[\mathbb{1}_{\{\tau^*<\tau_m\}}\Big(h(m) - p\theta\Big)\Big] + \mathsf{P}_x\big(\tau^* \geq \tau_m\big) \\ =& h(m) - p\theta - \mathsf{E}_x\Big[\mathbb{1}_{\{\tau^*\geq\tau_m\}}\Big(h(m) - p\theta\Big)\Big] + \mathsf{P}_x\big(\tau^* \geq \tau_m\big) \\ \leq& h(m) - p\theta + (p\theta + 1)\mathsf{P}_x\big(\tau_m \leq T\big). \end{aligned}$$

Letting $x \uparrow \infty$ we have $\mathsf{P}_x\big(\tau_m \leq T\big) \downarrow 0$. Then, letting $m \uparrow \infty$ we also have $h(m) \downarrow \alpha\gamma$, so that

$$\lim_{x\to\infty} v(0,x) \leq \alpha\gamma - p\theta.$$

The latter contradicts $v(0, x) \geq h(x) \geq \alpha\gamma$ for all $x \in \mathbb{R}_+$ (recall (2.15)). Hence, it cannot be $c(x) \geq \theta$ on $[\hat{x}_2, \infty)$.

Now we prove that indeed it cannot be $c(x) > 0$ on $[\hat{x}_2, \infty)$. By way of contradiction, assume the latter holds. For $\theta > 0$ let us introduce the auxiliary problem with value function

$$v_\theta(t,x) = \sup_{0\leq\tau\leq T+\theta-t} \mathsf{E}_x\Big[h(X_{\tau\wedge\tau^\dagger}) - \int_0^{\tau\wedge\tau^\dagger}\Big(p + qe^{-X_s}\Big)\mathrm{d}s\Big].$$

Since h and the process X are time-homogeneous we clearly have $v_\theta(t, x) = v(t - \theta, x)$ for all $t \in [\theta, T]$ and $x \in \mathbb{R}_+$. In particular, since we are assuming $c(x) > 0$ for $x \geq \hat{x}_2$, the optimal stopping boundary for the auxiliary problem is $c_\theta(x) = \theta + c(x)$ for $x \geq \hat{x}_2$. Hence, $c_\theta(x) \geq \theta$ for all $x \geq \hat{x}_2$. Then, by the same argument above with $v_\theta(0, x)$ instead of $v(0, x)$ we reach again a contradiction. □

Case (i) can be formally interpreted as the limiting situation of Case (ii) with $\hat{x}_1 = \infty$. By the same arguments above we obtain the analogue of (i) and (ii) in Proposition 5.3 but with $c(x) = 0$ for $x \geq x_2$ therein and $c(x_2-) = 0$. Once again continuity of the boundary follows from Theorem 5.11.

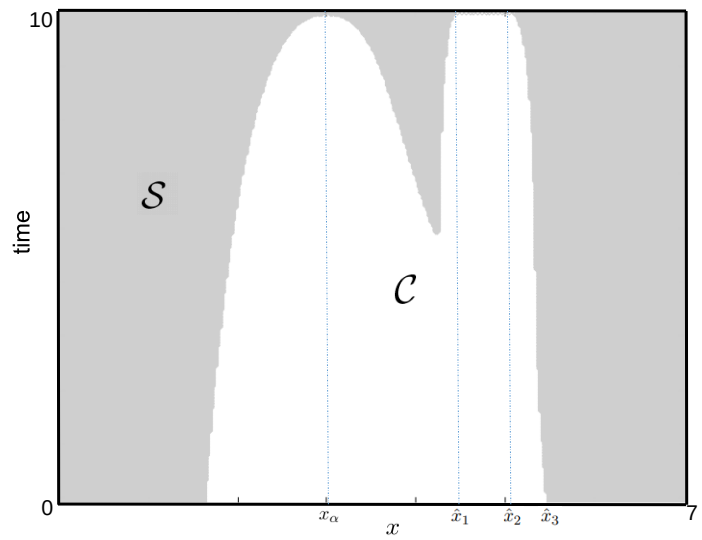

FIGURE 4. The optimal surrender region and boundary in Case (ii) of management fees (see Proposition 6.1).

The presence of management fees paid at a constant (proportional) rate, reduces the incentive created by the bonus mechanisms in the policy. That is why we observe an upper exercise boundary at which the policyholder surrenders the contract when the expected gains from staying in the contract (and collecting the bonus rate) are outweighed by the expected cost of future management fees.

APPENDIX

Proof of Lemma 5.1. Here we borrow arguments from the proof of [13, Thm. 1]. To keep a simple notation, in what follows we set $X_t = X_t^{x_\alpha}$ everywhere. From the equality

$$|X_t - x_\alpha| = \int_0^t \operatorname{sign}(X_s - x_\alpha)\mathrm{d}X_s + L_t^{x_\alpha},$$

we deduce that

$$\begin{aligned} L_{s\wedge\rho_\varepsilon}^{x_\alpha} &= |X_{s\wedge\rho_\varepsilon} - x_\alpha| - \int_0^{s\wedge\rho_\varepsilon} \operatorname{sign}(X_u - x_\alpha)\mathrm{d}X_u \\ &= |X_{s\wedge\rho_\varepsilon} - x_\alpha| - \int_0^{s\wedge\rho_\varepsilon} \operatorname{sign}(X_u - x_\alpha)\sigma\mathrm{d}W_u \\ &\qquad - \int_0^{s\wedge\rho_\varepsilon} \operatorname{sign}(X_u - x_\alpha)\pi(X_u)\mathrm{d}u. \end{aligned}$$

Notice that $\pi(\cdot)$ is bounded on $[x_\alpha - \varepsilon, x_\alpha + \varepsilon]$ by a constant $c_{\pi,\varepsilon} > 0$. Moreover, since $|X_{s\wedge\rho_\varepsilon} - x_\alpha| \leq \varepsilon$, then $1 \geq \varepsilon^{-p}|X_{s\wedge\rho_\varepsilon} - x_\alpha|^p$ for $p > 0$. Taking expectation and using these two observations we get

$$\begin{aligned} \mathsf{E}_{x_\alpha}\left[L_{s\wedge\rho_\varepsilon}^{x_\alpha}\right] &= \mathsf{E}_{x_\alpha}\left[|X_{s\wedge\rho_\varepsilon} - x_\alpha|\right] - \mathsf{E}_{x_\alpha}\left[\int_0^{s\wedge\rho_\varepsilon} \operatorname{sign}(X_u - x_\alpha)\,\pi(X_u)\mathrm{d}u\right] \\ &\geq \mathsf{E}_{x_\alpha}\left[|X_{s\wedge\rho_\varepsilon} - x_\alpha|\right] - \ c_{\pi,\varepsilon}\mathsf{E}_{x_\alpha}[s\wedge\rho_\varepsilon] \\ &\geq \varepsilon^{-p}\,\mathsf{E}_{x_\alpha}\left[|X_{s\wedge\rho_\varepsilon} - x_\alpha|^{1+p}\right] - c_{\pi,\varepsilon}\ s. \end{aligned}$$

The expectation of the absolute value above can be estimated using the integral form of the SDE for X and the inequality $|a+b|^{1+p} \geq \frac{1}{2^{1+p}}|a|^{1+p} - |b|^{1+p}$, for all $a, b \in \mathbb{R}$ (see, [25, Ch. 8, Sec. 50, p. 83]). That is,

$$
\begin{aligned}
&\mathsf{E}_{x_\alpha}\left[L^{x_\alpha}_{s\wedge\rho_\varepsilon}\right] \\
&\geq \varepsilon^{-p}\,\mathsf{E}_{x_\alpha}\left[\left|\int_0^{s\wedge\rho_\varepsilon}\pi(X_u)\mathrm{d}u + \sigma W_{s\wedge\rho_\varepsilon}\right|^{1+p}\right] - \; c_{\pi,\varepsilon}\; s \\
&\geq \varepsilon^{-p}\left\{\left(\frac{\sigma}{2}\right)^{1+p}\mathsf{E}_{x_\alpha}\left[|W_{s\wedge\rho_\varepsilon}|^{1+p}\right] - \mathsf{E}_{x_\alpha}\left[\left|\int_0^{s\wedge\rho_\varepsilon}\pi(X_u)\mathrm{d}u\right|^{1+p}\right]\right\} - \; c_{\pi,\varepsilon}\; s \qquad (6.2)\\
&\geq \varepsilon^{-p}\left\{\left(\frac{\sigma}{2}\right)^{1+p}\mathsf{E}_{x_\alpha}\left[|W_{s\wedge\rho_\varepsilon}|^{1+p}\right] - c^{1+p}_{\pi,\varepsilon}\; s^{1+p}\right\} - \; c_{\pi,\varepsilon}\; s.
\end{aligned}
$$

Now Burkholder-Davis-Gundy inequality and Doob's inequality imply that there exists a positive constant c_p such that

$$
\begin{aligned}
\mathsf{E}_{x_\alpha}\left[|W_{s\wedge\rho_\varepsilon}|^{1+p}\right] &\geq c_p\,\mathsf{E}_{x_\alpha}\left[(s\wedge\rho_\varepsilon)^{\frac{1+p}{2}}\right] \geq c_p\,\mathsf{E}_{x_\alpha}\left[\mathbb{1}_{\{s<\rho_\varepsilon\}}\right] s^{\frac{1+p}{2}} \\
&= c_p\, s^{\frac{1+p}{2}}\big(1 - \mathsf{P}_{x_\alpha}(\rho_\varepsilon \leq s)\big).
\end{aligned}
$$

Inserting the last inequality in (6.2) we get

$$
\mathsf{E}_{x_\alpha}\left[L^{x_\alpha}_{s\wedge\rho_\varepsilon}\right] \geq c_1\; s^{\frac{1+p}{2}} - c_2\left(s + s^{1+p} + \mathsf{P}_{x_\alpha}\left(\rho_\varepsilon \leq s\right) s^{\frac{1+p}{2}}\right) =: \lambda(s),
$$

for some suitable positive constants $c_1 = c_1(\varepsilon, p)$ and $c_2 = c_2(\varepsilon, p)$. Since $p \in (0,1)$, in the limit as $s \downarrow 0$ we get

$$
\lambda(s) = c_1 s^{\frac{1+p}{2}}\left[1 - \tfrac{c_2}{c_1}\big(\mathsf{P}_{x_\alpha}\left(\rho_\varepsilon \leq s\right) + s^{\frac{1-p}{2}} + s^{\frac{1+p}{2}}\big)\right] \geq \tfrac{1}{2}c_1 s^{\frac{1+p}{2}}
$$

because $\mathsf{P}_{x_\alpha}\left(\rho_\varepsilon \leq s\right) + s^{\frac{1-p}{2}} + s^{\frac{1+p}{2}} \downarrow 0$ as $s \to 0$. Given that $\frac{1+p}{2} \in (\frac{1}{2}, 1)$ we have $\lambda(s) > \ell \cdot s$ as $s \downarrow 0$ for any constant $\ell > 0$. That implies the claim of the lemma. □

Acknowledgment: The authors were financially supported by Sapienza University of Rome, research project "*Life market: a renewal boost for quantitative management of longevity and lapse risks*", grant no. RM11916B8953F292. T. De Angelis gratefully acknowledges support via EPSRC grant EP/R021201/1, "*A probabilistic toolkit to study regularity of free boundaries in stochastic optimal control*".

References

[1] Andreatta, G., Corradin, S., Valuing the Surrender Options Embedded in a Portfolio of Italian Life Guaranteed Participating Policies: a Least Square Monte Carlo Approach, in *Real Option theory meets practice, 8th Annual International Conference*, Montréal Canada, June 17-19, 2004.

[2] Bacinello, A.R., Fair Valuation of a Guaranteed Life insurance Participating Contract Embedding a Surrender Option, *J. Risk and Insurance*, 70 (2003), pp. 461-487.

[3] Bacinello, A.R., Biffis, E., Millossovich, P., Pricing Life Insurance Contracts with Early Exercise Features, *J. Computational Appl. Math.*, 233 (2009), pp. 27-35.

[4] Bass, R.F., *Diffusions and elliptic operators*, Springer-Verlag, New York 1998.

[5] Bayraktar, E., Xing, H., Analysis of the optimal exercise boundary of American options for jump diffusions, *SIAM J. Math. Anal.*, 41(2) (2009), pp. 825-860.

[6] Bensoussan, A., Lions, J.L., *Applications of variational inequalities in stochastic control*, Studies in Mathematics and its Applications, Vol. 12, North-Holland 1982.

[7] Cannon, J.R., *The one-dimensional heat equation* (No. 23), Cambridge University Press, 1984.

[8] Chen, X., Chadam, J., A mathematical analysis of the optimal exercise boundary for American put options. *SIAM J. Math. Anal.*, 38(5) (2007), pp. 1613-1641.

[9] Cheng, C., Li, J., Early default risk and surrender risk: Impacts on participating life insurance Policies, *Insurance Math. Econom.*, 78, (2018), pp. 30-43.
[10] Chu, C.C., Kwok, Y.K., Pricing Participating Policies with Rate Guarantees, *Int. J. Theor. Appl. Finance*, 9, No. 4 (2006), pp. 517-532.
[11] Dai, M., Yi, F., Finite-horizon optimal investment with transaction costs: a parabolic double obstacle problem, *J. Diff. Eq.*, 246(4) (2009), pp. 1445-1469.
[12] De Angelis, T., A note on the continuity of free-boundaries in finite-horizon optimal stopping problems for one dimensional diffusions, *SIAM J. Control Optim.*, 53 (1) (2015), pp. 167-184.
[13] De Angelis, T., Kitapbayev, Y., On the optimal exercise boundaries of swing put options, *Math. Oper. Res.* 43 (1) (2018), pp. 252-274.
[14] De Angelis, T., Peskir, G., Global C^1 regularity of the value function in optimal stopping problems, *Ann. Appl. Probab.* 30 (3) (2020), pp. 1007-1031.
[15] De Angelis, T., Stabile, G. On the free boundary of an annuity purchase, *Finance Stoch.* 23 (2019), pp. 97-137.
[16] Du Toit, J., Peskir, G., The trap of complacency in predicting the maximum, *Ann. Appl. Probab.* 35 (1) (2007), pp. 340-365.
[17] Ekström, E., Vaicenavicius, J., Optimal stopping of a Brownian bridge with an unknown pinning point, *Stoch. Process. Appl.* 130 (2) (2020), pp. 806-823.
[18] Fard, A.F., Siu, T.K., Pricing participating products with Markov-modulated jump-diffusion process: An efficient numerical PIDE approach, *Insurance Math. Econom.*, 53 (2013), pp. 712-721.
[19] Friedman, A., *Partial differential equations of parabolic type*, Englewood Cliffs, N.J., Prentice Hall, 1964.
[20] Friedman, A., Parabolic variational inequalities in one space dimension and smoothness of the free boundary, J. Funct. Anal., 18(2) (1975), pp. 151-176.
[21] Friedman, A., *Variational principles and free-boundary problems*, John Wiley & Sons, Inc., New York, 1982.
[22] Grosen, A., Jørgensen, P. L., Fair Valuation of Life Insurance Liabilities. The Impact of Interest Rate Guarantees, Surrender Options, and Bonus Policies, *Insurance Math. Econom.*, 26 (2000), pp. 37-57.
[23] Itô, K., McKean, H.P., *Diffusion processes and their sample paths*. Springer, Berlin, 1965.
[24] Karatzas, I., Shreve, S., *Brownian Motion and Stochastic Calculus*, Second Edition Springer, 1991.
[25] Kolmogorov, A.N., Fomin, S.V., *Elements of the theory of functions and functional analysis*, vol. 1, Dover Publications, Mineola, NY, 1999.
[26] Ladyzenskaja, O.A., Solonnikov, V.A., Ural'ceva, N.N., *Linear and quasilinear equations of parabolic type*, (Translated from Russian) by Transl. Math. Monogr. vol. 23,. AMS, Providence, RI, 1967.
[27] Lions, P.L., Sznitman, A.S., Stochastic differential equations with reflecting boundary conditions, *Comm. Pure Appl. Math.* **37** (1984), pp. 511-537.
[28] Nelson, D.B., Ramaswamy, K., Simple Binomial Processes as Diffusion Approximations in Financial Models, *Rev. Financial Studies*, 3 (1990), pp. 393-430.
[29] Nolte, S., Schneider, J.C., Don't Lapse into temptation: a behavioral explanation for policy surrender, *J. Banking and Finance*, 79 (2017), pp. 12-27.
[30] Peskir, G., Shiryaev, A.N., *Optimal Stopping and Free-Boundary Problems*, Lectures in Mathematics, ETH Zürich, Birkhäuser (2006).
[31] Protter, P.E., *Stochastic integration and differential equations*, second edition. Springer-Verlag, Berlin, Heidelberg, New York (2004).
[32] Rogers, L.C.G., Williams, D., *Diffusions, Markov Processes and Martingales*, Vol. 2, second edition. Cambridge University Press 2000.
[33] Shanahan, B., Alavi Fard, F. and van der Hoek, J., Pricing participating policies under the Meixner process and stochastic volatility, *Scand. Actuar. J.*, 7 (2017), pp. 559-583.
[34] Siu, T.K., Fair Valuation of Participating Policies with Surrender Options and Regime Switching, *Insurance Math. Econom.*, 37 (2005), pp. 533-552.
[35] Stabile, G., *Optimal timing of the annuity purchase: a combined stochastic control and optimal stopping problem*, Int. J. Theor. Appl. Finance, **9** (2) (2006), pp. 151-170.

M.B. Chiarolla: Dipartimento di Scienze dell'Economia, Centro Ecotekne, Università del Salento, Lecce, Italy.
Email address: maria.chiarolla@unisalento.it

T. De Angelis: School of Management and Economics (Dept. ESOMAS) University of Torino and Collegio Carlo Alberto, Torino, Italy.
Email address: tiziano.deangelis@unito.it

G. Stabile: Dipartimento di Metodi e Modelli per l'Economia, il Territorio e la Finanza, Sapienza-Università di Roma, Roma, Italy
Email address: gabriele.stabile@uniroma1.it